PB83-253492

Airborne Asbestos Levels in Schools: Design Study

Research Triangle Inst.
Research Triangle Park, NC

Prepared for

Environmental Protection Agency, Washington, DC

Jul 81

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA 560/5-81-006	2.
3. RECIPIENT'S ACCESSION NO. PB83 253492	
4. TITLE AND SUBTITLE Airborne Asbestos Levels in Schools: Design Study	5. REPORT DATE July 1981
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Bert Price, Donna Watts, Everett Logue, Tyler Hartwell	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Triangle Institute Post Office Box 12194 Research Triangle Park, North Carolina 27709	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Protection Agency Office of Pesticides and toxic Substances Washington, D.C. 20460	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE
15. SUPPLEMENTARY NOTES	

16. ABSTRACT

This document describes a proposed field study to collect data in schools that are to be used to analyze and validate two asbestos exposure assessment algorithms as compared to levels of airborne asbestos. This field study would involve algorithm scoring (including bulk asbestos sampling) and air sampling in sites (e.g., classrooms) within selected schools. The objective of the planning study described in this report is to establish the characteristics of various alternative statistical designs (e.g., number and characteristics of sample sites) for the proposed field study and to recommend the most appropriate design.

The report is intended to provide EPA with an assessment of precision and completeness that can be expected from the data collected in the field study.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Airborne Asbestos Levels Asbestos in Bulk Samples Asbestos Algorithm Scoring Asbestos Levels in Schools Statistical Design		

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES 95
	20. SECURITY CLASS *(This page)*	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

United States
Environmental Protection
Agency

Toxic Substances
Washington DC 20460

EPA-560/5-81-006
July 1981

PB83-253492

Toxic Substances

EPA

Airborne Asbestos Levels in Schools: Design Study

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

EPA Report Number 560/5-81-006
July 1981

AIRBORNE ASBESTOS LEVELS IN SCHOOLS: DESIGN STUDY

by

B. Price
Battelle Memorial Institute
Columbus, Ohio 43201

and

D. Watts
E. Logue
T. Hartwell
Research Triangle Institute
Research Triangle Park, North Carolina 27709

EPA Contract Number 68-01-5848
RTI Project Number 1864-23

EPA Task Manager: Cindy Stroup
EPA Project Officer: Joe Carra

Prepared for:

Design and Development Branch
Exposure Evaluation Division
Office of Pesticides and Toxic Substances
U.S. Environmental Protection Agency
Washington, D.C.

DISCLAIMER

This report was prepared under contract to an agency of the United States Government. Neither the United States Government nor any of its employees, contractors, subcontractors, or their employees makes any warranty, expressed of implied, or assumes any legal liability or responsibility for any third party's use or the results of such use of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe on privately owned rights.

Publication of the data in this document does not signify that the contents necessarily reflect the joint or separate views and policies of each sponsoring agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

		Page
I.	EXECUTIVE SUMMARY	1
II.	INTRODUCTION/BACKGROUND	5
III.	THE VALIDATION PROBLEM	9
	A. Validation: Conceptual Background	9
	1. Role of Air Sampling	10
	2. Framework	11
	B. Validation: Specific	14
	1. Description of the Data	18
	2. Validation for SA	19
	3. Validation for DTA	20
IV.	SAMPLE SIZE AND PRECISION	23
	A. Sources of Variation	23
	B. Sample Size: SA: Controlled Experiment	25
	1. SA Formulation 1	25
	2. SA Formulation 1: Summary	34
	3. SA Formulation 2	36
	4. Comparison: Formulation 1 versus Formulation 2	45
	C. Sample Size: Decision Tree Algorithm (DTA): Controlled Experiment	48
	D. Sample Size: Measurement Error	51
	1. SA Formulation 2	52
	2. DTA	53
V.	DESCRIPTION OF AVAILABLE SITES IN MONTGOMERY COUNTY AND NEW YORK CITY	59
	A. The Montgomery County Data	59
	B. The New York City Data	59
VI.	SAMPLE DESIGN	65
	A. Introduction	65
	B. Study Area	65
	C. Overview of the Sample Design	66
	D. Construction and Stratification of the First-Stage Frame	67
	E. Allocation and Selection of the First-Stage Sample	71
	F. Construction and Stratification of the Second-Stage Frame	74
	G. Allocation and Selection of the Second-Stage Sample	76
	H. Data Collection at Sample Sites	77
	I. Design Effect	78
	J. Sampling Strategy	82
REFERENCES		84
APPENDIX A:	The Noncentral t Distribution	

LIST OF TABLES

Number	Title	Page
1	Factor Weights for SA Score	13
2	Site Description for DTA	16
3	Required Sample Size (n) for Testing H_0: $\mu_{10} = \mu_B$ vs. H_1: $\mu_{10} - \mu_B = K\sigma$ (n sites in each group)	28
4	Probability That $Y_{10} \leq Y_B$ When Testing H_0: $\mu_{10} = \mu_B$ versus H_1: $\mu_{10} - \mu_B = K\sigma$ ...	30
5	Raw Meansurement Units (ng/m^3) For The Alternative Hypothesis When Testing H_0: $\mu_{10} = \mu_B$ versus H_1: $\mu_{10} - \mu_B = K\sigma$ ...	32
6	Effect of Increased Standard Deviation On Raw Measurement Units (ng/m^3) For The Alternative Hypothesis When Testing H_0: $\mu_{10} - \mu_B$ versus H_1: $\mu_{10} - \mu_B = K\sigma$ ...	33
7	Summary of Power (%) For Testing H_o: $\mu_{10} - \mu_B$ versus H_1: $\mu_{10} - \mu_B = K\sigma$ ...	35
8	Full Factorial Sampling Plan	39
9	Full-Factorial Sampling Plan, Average Factor Scores	40
10	Formulation 2: Full-Factorial Design H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/\Delta$, n = 32, Degrees of Freedom = 30, $\sum_S = 138.8$; Body of Tables Gives Power of the Test (%)	41
11	Formulation 2: Half-Fraction Design H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/\Delta$, n = 16, Degrees of Freedom = 14, $\sum_S = 98.2$; Body of Table Gives Power of the Test (%)	42
12	Formulation 2: Optimal Design H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/30$ (i.e., Δ=30), $\sum_S = 49.9 \times n^{1/2}$ Body of the Table Gives Power of the Test (%)	44

List of Tables (cont'd)

Number	Title	Page
13	Comparison: Formulation 1 Versus Formulation 2. $\alpha = .05$, $\Delta = 30$; Body of Table Presents Power (%)	47
14	Statistical Test: DTA; H_0: $\mu_A = \mu_{DA}$; H_1: $\mu_A - \mu_{DA} = K\sigma$, Body of Table Presents Power of Test (%)	50
15	DTA: H_0: $\eta_A = \eta_{DA}$; H_1: $\eta_A - \eta_{DA} = (1-2p)K\sigma$ Body of the Table Gives Power of Degradation When Sites Are Misclassified. Significance Level = .05	56
16	Distribution of Montgomery County Sites ..	60
17	Distribution of New York Sites	62
18	Asbestos Content of New York City Public Schools	69
19	Distribution of Asbestos-Containing Public Schools in New York City with Respect to Asbestos Content, Friability, Condition, Exposure, and Accessibility	70
20	First-Stage Strata	72
21	Distribution of New York City Public Schools by Selected Size Categories	73
22	Second-Stage Strata	75
23	The Clustering Effect Corresponding to Selected Values of ρ and n_2	80

LIST OF FIGURES

Number	Title	Page
1	Relationship Between Air Levels and SA Scores.....	12
2	Decision Tree Algorithm (DTA)	15
3	Hypothetical Air Levels for DTA Groups.....	17
4	Some Effects of an Erroneous Linear Assumption.....	46

ACKNOWLEDGMENTS

The authors gratefully acknowledge the helpful suggestions of Cindy Stroup, Joe Breen and Larry Longanecker of the U.S. Environmental Protection Agency, Washington, D.C., and Steve Williams of the Research Triangle Instititue, Research Triangle Park, North Carolina.

I. EXECUTIVE SUMMARY

A field study has been proposed to collect data in schools that are to be used to analyze and validate two asbestos exposure assessment algorithms as compared to levels of airborne asbestos. This field study would involve algorithm scoring (including bulk asbestos sampling) and air sampling in sites (e.g., classrooms) within selected schools. The objective of the planning study described in this report is to establish the characteristics of various alternative statistical designs (e.g., number and characteristics of sample sites) for the proposed field study and to recommend the most appropriate design.

The report is intended to provide EPA with an assessment of precision and completeness that can be expected from the data collected in the field study.

The approach is to formulate various operational definitions of validation and to develop the statistical characteristics of specific designs with respect to these definitions. In order to compare designs, it is necessary to incorporate assumptions that simplify the validation problem. Although the problem is simplified, the comparisons of statistical designs are still informative. The major results of the investigation follow.

(1) Using information based on data that are currently available, it has been determined that between 32 and 64 sites are required in the field study in order to meet most reasonable statistical objectives. These sites would be selected in approximately 16 schools.

(2) The basic 32 sites and 16 schools should be selected according to the two-stage sampling design with stratification imposed at each stage as described in the report. The two-stage sampling procedure would base selections of sites and schools on the scores of the factors that make up the as-

sessment algorithms (e.g., asbestos content, friability, etc.). The first stage of sampling would select schools and the second stage would select the sites within the schools.

(3) Based on currently available asbestos assessment algorithm data, New York City schools are recommended as the location for the field study if local cooperation is forthcoming.

(4) It is recommended that the initial 3 or 4 sites selected for air sampling for the proposed study have algorithm scores that are at the high and low ends of the algorithm scoring scale (for example, two sites with high asbestos, friability, exposure, accessibility, and bad condition; and two sites with low asbestos, friability, exposure, accessibility, and good condition). Air sampling results from these initial sites would then be compared (as well as results from background samples at each site) to determine if clear differences are evident. This analysis could be used to determine if additional air sampling is warranted.

(5) The variabilities found at a fixed site in exposure assessment algorithm scoring and in airborne asbestos concentration levels are the key factors in determining sample size. These parameters should be monitoried in the early stages of the field study to determine that they have values that are consistent with the values used to develop the sampling plan. A sample size adjustment upward from 32 to 64 sites may be necessary if the variability encountered is larger than expected.

Specific rules for selecting sites and collecting measurements are presented in the body of the report. In addi-

tion, a description is presented of the population of potential schools and sites in two school districts (Montgomery County, Ohio, and New York City) where algorithm data are currently available.

II. INTRODUCTION/BACKGROUND

In March 1979, EPA initiated a rulemaking proceeding under the Toxic Substances Control Act (TSCA) regarding friable asbestos-containing materials in schools, (44 FR 177790). In September 1979, an Advance Notice of Proposed Rulemaking (ANPRM) was published (44 FR 54676) that discussed EPA's plan for rulemaking. EPA's plan includes "(1) requiring surveys of schools to determine whether they contain friable asbestos-containing materials, requiring that an exposure assessment be performed for all such materials identified, and requiring that friable asbestos-containing materials be marked; (2) requiring corrective actions with respect to friable asbestos-containing materials for which the exposure assessment exceeds a level determined by EPA as presenting an unreasonable risk; and (3) requiring periodic reevaluation of the friable asbestos-containing materials to determine whether the exposure assessment is still valid or whether additional corrective action is required under the regulation."

Upon confirmation that friable material is present and that the material contains at least 1 percent asbestos, an exposure assessment would be required. A numerical based inspection scheme (also referred to as algorithm) has been proposed to serve as the exposure assessment tool. The algorithm value is determined from scores assigned to eight factors--condition, water damage, exposure, accessibility, activity, presence of an air plenum, percent asbestos content, and friability. The factor scores are summarized to form an exposure number which is compared to a preestablished numerical scale that indicates what correction or control, if any, is necessary.

Two forms of summarization have been proposed--Sawyer's Algorithm (SA) and the Decision Tree Algorithm (DTA). The usefulness of either of these summarizations in rulemaking has been an ongoing concern to EPA. Over the past 2 years,

Preceding page blank

there have been a number of investigations directed at validating various algorithms (see Patton et al. 1980; Price and Townley 1980a; Logue and Hartwell 1981; EPA 1980a). The experience gained has resulted in the design of a field study using air sampling to validate SA and DTA (Price et al. 1980b). The recommended field program for collecting data to validate these algorithms consists of a set of activities that must be carefully controlled if the resulting data are to be informative. Care must be taken to insure (1) that the technical and logistical problems associated with field air monitoring are satisfactorily solved, and (2) that the statistical sampling plan provides appropriate levels of precision to support the types of regulatory decisions that are being considered.

The research presented in this report addresses the statistical issue--Item (2) above. Typically, not enough effort is devoted to understanding the types of precision statements that are defensible and consistent with a proposed sampling plan. There is often a "gap" between the precision and completeness of the information expected by the regulator, and the precision and completeness that can be validly associated with the data and the statistical analysis. The consequences of not closing this "gap" are (1) investments in research that are large, but not large enough to provide a usable research result, and (2) investments that should not have been made because there was no reasonable agreement on the level of precision to be expected and therefore no chance that the research objective would be met. The goal of the current planning study is to close the "expectations gap" in order to provide EPA with an opportunity to reevaluate the resources required to meet the resources required to meet the objectives of the suggested validation study.

This report includes an evaluation of alternative statistical plans that have been suggested for the field study.

Operational definitions of "validity" are discussed. The two measurement processes--air sampling and algorithm scoring--are characterized in terms of sources of variability. The relationship between sample size and precision is developed for the statistical sampling plans that have been proposed.

III. THE VALIDATION PROBLEM

EPA's objective is to select a numerical based visual inspection scheme to serve as an exposure assessment tool. One important aspect of the selection process is validation --that is, documenting that the assessment tool operates as intended. At this time, there are two proposed tools for exposure assessment: Sawyer's Algorithm (SA) and the Decision Tree Algorithm (DTA). Operational definitions of validity are closely linked to the particular tool that is being validated. In this section of the report, we present alternative definitions of validity that are used later in the analysis of sample size, statistical design, and precision.

A. Validation: Conceptual Background

The primary concern in assessing exposure to asbestos in schools is to be able to initiate corrective action in those situations that present an unreasonable risk to the health of persons who use school buildings. The health effects associated with exposure to asbestos fibers have been analyzed using epidemiologic and experimental data. It has been shown that exposure to asbestos via inhalation increases the risk of numerous diseases including asbestiosis pleural and peritoneal mesothelioma, and cancers of the lung and other organs (45 FR 61966, Technical Support Document for Regulatory Action Against Friable Asbestos-Containing Materials in School Buildings, USEPA, August 1980). As formulated, validation of the proposed exposure assessment methods involves two relationships:

(1) the relationship of the risk of adverse health effects to dose levels of airborne asbestos fibers; and

(2) the relationship between levels of airborne asbestos fibers and scores on the specific exposure assessment algorithm under consideration.

The scope of the current report is restricted to validation of the second relationship. Validity of the dose-response

Preceding page blank

relationship is assumed. (See 45 FR 61969 and the associated list of support documents, 45 FR 61980-61986.)

1. Role of Air Sampling

It has been suggested that exposure assessment should be directly based on measurements of airborne asbestos concentration levels rather than using a proxy such as a visual inspection scoring algorithm. Direct measurement would be accomplished by air sampling, and the air sampling approach has been considered. However, "EPA has determined that air sampling is an inappropriate test to determine if an asbestos exposure problem exists in school buildings" (45 FR 61978). There are at least four objections. First, air sampling measures the concentration of airborne asbestos fibers only at the time the sample is taken. However, airborne concentration levels are the result of an episodic process more than a continuous process. Air sampling over relatively short time periods (as would be necessary in an assessment program) could easily lead to misclassification of a potentially hazardous site. Second, the exposure assessment is supposed to identify not only currently hazardous sites, but it is also supposed to provide information about sites that may become hazardous in the future. Air sampling is capable of identifying a current problem only. Third, it is not feasible to use air sampling in an extensive assessment program because of seemingly insurmountable technical and logistical problems. To be effective, air sampling would have to be conducted for periods of at least one week at sites where normal activity was in progress. Difficulties of implementation include the nuisance factor of sampling equipment operating in a school environment, problems in establishing background asbestos levels, and assuring the integrity of the samples collected in an uncontrolled field setting. Finally, the resources necessary for the collection and analysis of scientifically valid air samples are not available. Although these diffi-

culties render air sampling inappropriate for widespread use as an exposure assessment tool, air sampling may be used as a basis for validating the proposed assessment methods.

For the implementation of the validation study, it has been agreed that airborne asbestos data will only be used to validate the algorithm factors that are indicators of current asbestos levels. Validation of those factors that are mainly predictors of future health risk must be established by other means. The validation study is designed to collect air data over time intervals that are long enough to capture the effects of an episodic process. Air sampling will be conducted for one week during regular school hours so that regularly occurring activity is represented. These long sampling periods are acceptable for the validation study because the overall scope of sampling is limited. The related technical and logistical problems are manageable because the validation field program is small relative to the field program that would be necessary if air sampling were used as the exposure assessment method. Also, other potential problems such as the nuisance factors associated with classroom sampling and assuring the integrity of samples may be costly to solve, but they are manageable because the validation effort is well defined, and appropriate controls can be imposed.

2. Framework

As mentioned above, our concern is the validity of the decision rules related to taking a corrective action or deferring action to a later time. We now describe the framework underlying the decision rules for both SA and DTA.

When considering SA, the decision rule is derived from the relationship between air levels of asbestos and SA scores. Conceptually, the relationship takes the form shown in Figure 1. As SA increases, airborne asbestos concentration increases (see Table 1 which shows how SA is computed).

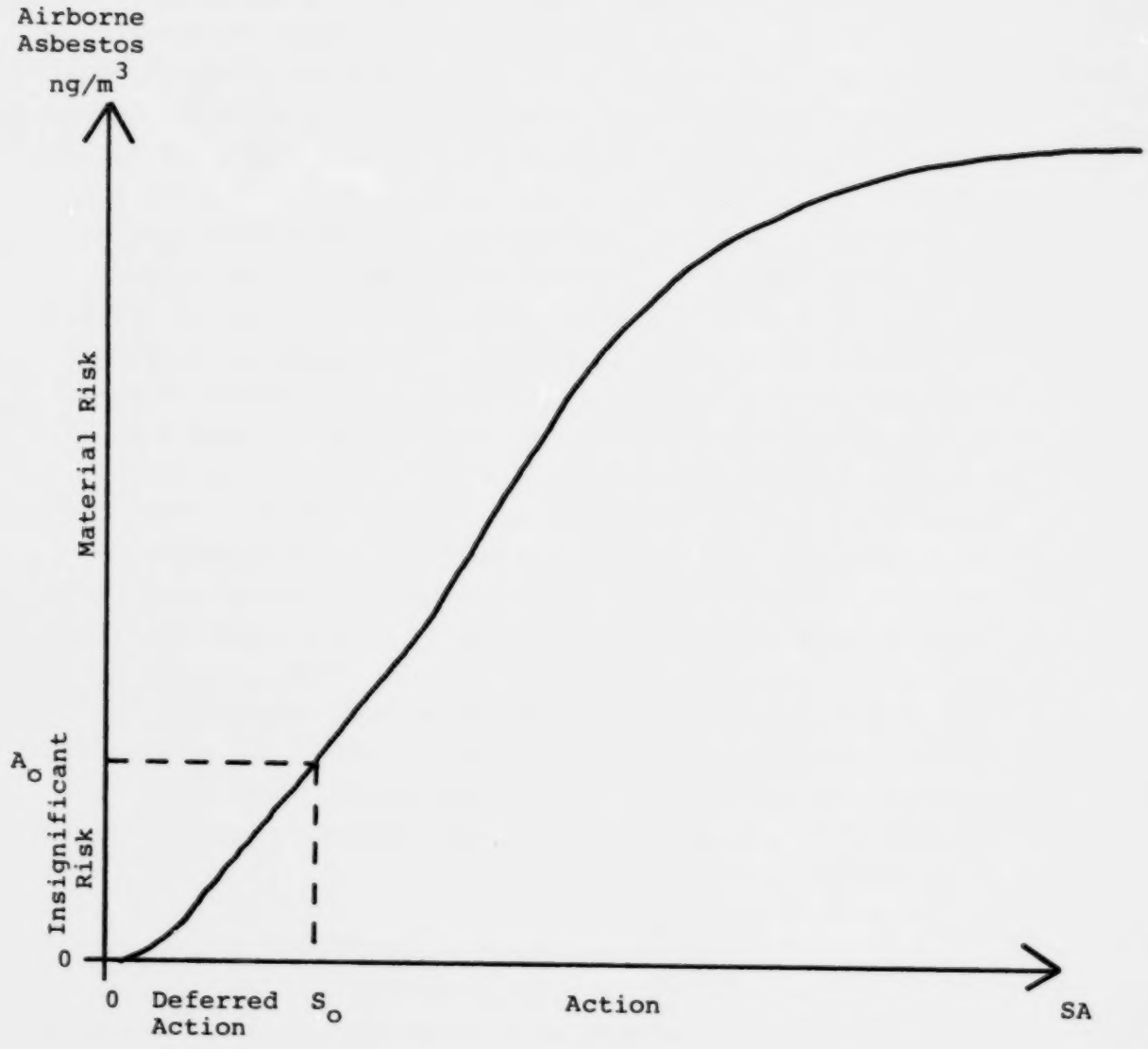


Figure 1. Relationship between air levels and SA scores.

Table 1. Factor Weights for SA Score

Factor	Weighted Scores
1. Condition (F1)	
No Damage	0
Moderate Damage	2
Severe Damage	5
2. Accessibility (F2)	
Not Accessible	0
Rarely Accessible	1
Accessible	3
3. Part of Air Moving System (F3)	
No	0
Yes	1
4. Exposure (F4)	
Material is not exposed	0
10 percent or less of the material is exposed	1
Greater than 10 percent of the material is exposed	4
5. Water Damage (F5)	
No water damage	0
Minor water damage	1
Moderate or major water damage	2
6. Activity or Movement (F6)	
None or low activity level	0
Moderate activity level	1
High activity level	2
7. Friability (F7)	
Not friable	0
Low friability	1
Moderate friability	2
High friability	3
8. Percentage Asbestos (F8)	
Less than or equal to 1 percent	0
Greater than 1 percent and less than or equal to 50 percent	2
Greater than 50 percent	3

$$\text{SA SCORE} = (F_1 + F_2 + F_3 + F_4 + F_5 + F_6) \times F_7 \times F_8$$

In the figure, A_o represents the airborne asbestos level that defines the cutoff between safe and unsafe levels. A_o in turn determines S_o on the SA scale which represents the cutoff point defining corrective action versus the deferral of corrective action. It is assumed that the value of A_o has been determined from the dose-response analysis relating airborne asbestos levels to health risk.* Validation of SA is based on the precision associated with the establishment of the curve in Figure 1.

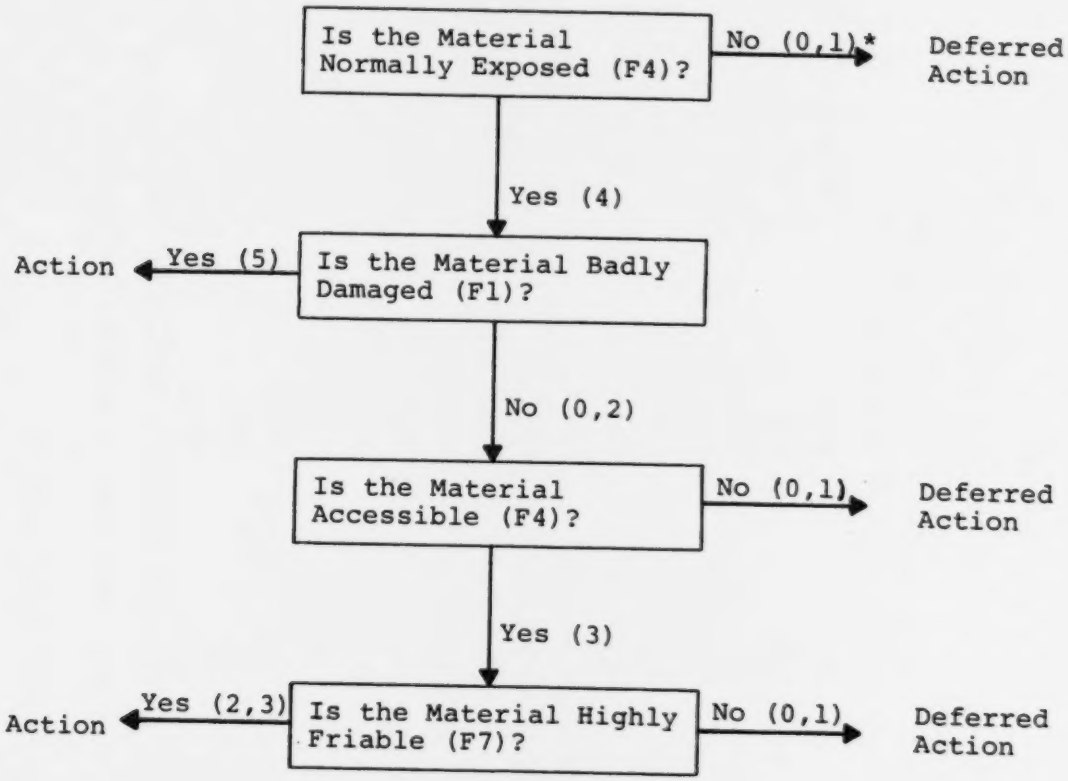
When considering DTA (Decision Tree Algorithm), the decision rule for action or deferred action arises from the classification of sites into one of two groups. For convenience, the groups are named Action (Group A) and Deferred Action (Group DA). A site is classified into Group A or Group DA according to the tree shown in Figure 2. Equivalently, there are 16 types of sites as shown in Table 2. Five sites belong to Group A; eleven sites belong to Group DA. From a conceptual perspective, there is a distribution of airbone asbestos concentrations corresponding to each group (see Figure 3). Validation of DTA involves the comparison of the characteristics of these distributions.

B. Validation: Specific

In the previous paragraphs, the conceptual framework for a validation analysis was developed. Before entertaining any questions concerning sampling design, sample size, and precision, it is necessary to formulate the validation problem in precise, quantitative terms. A few alternative formulations will be considered. Each formulation is specific and therefore each one appears to be limited in scope. However, it is necessary to focus on limited objectives if the ensuing discussion of sample size and precision is to be meaningful. The exercise is constructive in that

---

* As formulated, the validation analysis does not involve any specific values of A_o.



* The numbers in parentheses are factor weights (see Table 1).

NOTE: The decision tree is applied to sites with more than 1 percent asbestos in bulk sample.

Figure 2. Decision tree algorithm (DTA).

Table 2. Site Description for DTA[a]

Site ID	Factors[b]				Group
	Friability	Condition	Exposure	Accessibility	
2	Low	Good	Low	High	DA
3	Low	Good	High	Low	DA
5	Low	Bad	Low	Low	DA
8	Low	Bad	High	High	A
9	High	Good	Low	Low	DA
12	High	Good	High	High	A
14	High	Bad	Low	High	DA
15	High	Bad	High	Low	A
17	Low	Good	Low	Low	DA
20	Low	Good	High	High	DA
22	Low	Bad	Low	High	DA
23	Low	Bad	High	Low	A
26	High	Good	Low	Low	DA
27	High	Good	High	Low	DA
29	High	Bad	Low	Low	DA
32	High	Bad	High	High	A

a Decison tree is applied to sites with more than 1 percent asbestos in bulk sample.

b Friability (low) = 0 or 1 (see Table 1 for codes)
Condition (good) = 0 or 2
Exposure (low) = 0 or 1
Accessibility (high) = 3

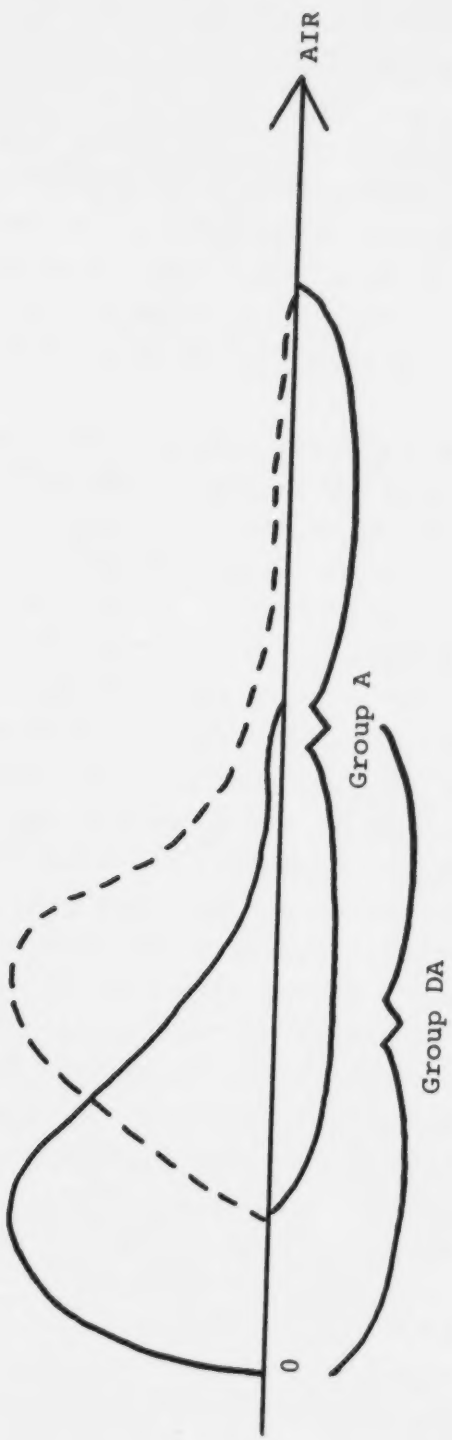


Figure 3. Hypothetical air levels for DTA groups.

it focuses attention on a precise formulation of the validation problem and allows for sensitivity analysis regarding sample size requirements.

1. Description of the Data

The first step in establishing a precise formulation of the validation problem is to characterize the type of data that will be collected. One set of assumptions about the data will serve each formulation of the validation problem whether it involves SA or DTA. The characterization follows.

It is planned that air sampling will be conducted at a flow rate of 5 liters/min., using 47mm Millipore filters mounted at a breathing height of 1.5 meters above the floor (see Price et al. 1980b). The samples will be analyzed by transmission electron microscopy (TEM). The basic measurement will be obtained by sampling for a full week of 40 hours when school activities are in progress.

Let us denote airborne asbestos concentration by X, measured in units of nanograms per cubic meter (ng/m^3). For a given site, let X_B and X_S denote the background levels and the concentration levels at the site, respectively. Background measurements will be taken near the site (e.g., outside the school). The exact location for background sampling will depend on the structure of the building and the location of potential sources of background asbestos. We shall use the conservative assumption that X_B and X_S are statistically independent and that $Y_B = \ln(X_B)$ and $Y_S = \ln(X_S)$ are normally distributed with expected value μ_B (μ_S) and common variance σ_B^2 ($\sigma_S^2 = \sigma_B^2$). The increment due to site is $Z = Y_S - Y_B$ which is normally distributed with the expected value $\eta_S = \mu_S - \mu_B$ and variance $\sigma^2 = 2\sigma_S^2$.

The variance, σ^2, is a composite of various components including analytical error, site variation, and sampling interval variation. As presented, σ^2, represents total vari-

ability. The roles of individual components of variability are discussed later in the report.

2. Validation for SA

We consider two formulations of the validation problem for SA.

Formulation 1:

Suppose that values of SA equal to 10 and 40 are used as decision cutoff points. In the discussion below, these two values are used to state specific statistical hypotheses to be used for validation. It is clear that the formulation applies as well to any two selected values of SA.

We expand the notation slightly to be able to indicate the level of SA at which the air data were collected. Let S take on the value of SA at the site under consideration. Consider two statistical hypotheses:

(a) $H_O: \mu_{10} = \mu_B$

$H_1: \mu_{10} > \mu_B$

(b) $H_O: \mu_{40} = \mu_{10}$

$H_1: \mu_{40} > \mu_{10}$

Hypothesis (a) tests whether or not measured asbestos concentration levels at sites where SA is 10 can be distinguished from background levels. Hypothesis (b) tests for differences in airborne asbestos levels between sites where SA is 40 and 10. This formulation of a validity test represents the minimum that would be required of the relationship between airborne asbestos and SA if decisions are to be based on scores of 10 and 40.

Formulation 2:

Assume that the relationship between Z and SA can be approximated by the model

$$Z = B_O + B_1 \cdot SA + \varepsilon$$

where ε has expected value equal to zero and standard deviation, σ. Test the hypothesis

$$H_0: B_1 = 0$$

$$H_1: B_1 > 0 .$$

This formulation of the validation test is an attempt to establish validity at every value of SA and includes hypotheses such as $H_0: \mu_{40} = \mu_{10}$ as a special case.

Comparison of these formulations regarding sample size and precision is found in a later section of this report.

3. Validation For DTA

The probability distribution and the parameters used to characterize the data for validating DTA are basically the same as used for SA. However, since the scores on SA are not part of the DTA scheme, a slightly different notation must be used.

Table 2 shows the 16 possible combinations that define unique sites based on the four factors that are used in DTA. Each site is classified into either the action group (Group A) or the deferred action group (Group DA) (see the right-hand column of Table 2). Denote the expected value of the air measurement for a given type of site by μ_i where i is the site ID. Then define

$$\mu_A = \sum_{\text{Group A}} \mu_i \; / \; 5$$

and

$$\mu_{DA} = \sum_{\text{Group DA}} \mu_i \; / \; 11$$

corresponding to Group A and Group DA, respectively. The statistical hypotheses are:

$$H_0: \mu_A = \mu_{DA}$$

$$H_1: \mu_A > \mu_{DA} .$$

This formulation of the validity test represents the minimum information that could be required on the relationship between airborne asbestos levels and DTA if DTA is to be used in rulemaking. Acceptance of H_1 indicates that it has been

confirmed that airborne asbestos levels are higher at sites designated for immediate corrective action than at sites where the decision is to defer corrective action.

IV. SAMPLE SIZE AND PRECISION

In this section, the relationship between sampling design, sample size, and precision is analyzed. The discussion is restricted to those designs that are presented in the Battelle study design report dated November 20, 1980. Validation has been formulated as a problem in statistical hypothesis testing (see previous section). In the testing problem, the sample size required depends on the size of the difference between the parameters of interest and the probability of detecting that difference. Precision enters through the variability associated with the process that is producing the measurements.* Both the magnitude and type of variation can be important. Variation is summarized in the parameter σ for the airborne asbestos process. In general, as variation increases, the probability of detecting any fixed difference declines.

As suggested earlier, variation may be a composite of many components--site, measurement, and model to name a few. Proper understanding and evaluation of these components are critical to the discussions of sample size that follow. The issue of variance components is discussed next. Then sample size tables are presented and discussed in relationship to alternative assumptions about the structure of σ^2.

A. Sources of Variation

There are two basic measurement processes to be considered--air sampling and algorithm scoring. Variation associated with the data collected from each process arises from a number of different sources. For air sampling, the variation may be due to:

(1) Site--Airborne asbestos levels vary because of systematic differences among sites. There is also a random component of variation that arises even when site characteristics are identical.

---

* In this discussion, precision may be thought of as the reciprocal of σ.

Preceding page blank

(2) Measurement error--The laboratory measurement procedure involving TEM to determine asbestos concentration is subject to error which can be characterized as random.

(3) Time--Collecting weekly samples at a given site under seemingly constant external conditions may result in a distribution of values when data from different weeks are pooled.

For algorithm scoring, the variation may be due to:

(1) Sites--Sites exhibit systematic differences.

(2) Rater--Different raters may score the same site differently. Also, any one rater may not be able to duplicate his or her score at a fixed site. This source of variation is similar to measurement error in laboratory procedures.

(3) Content--One important factor in the algorithm is based on the percentage of asbestos found in a bulk sample. Variability in content is the variability associated with the bulk sampling protocol. Included as sources of variation are

(a) location--random variation across locations where bulk samples are taken; and

(b) measurement error--random variation associated with the laboratory procedure.

The structure of variation imposed by the sources described above is used in the analysis of sample size that follows. Three cases are considered. The first case is developed from the most simplifying assumptions about sources of variation. That case corresponds to performing the validation study in a controlled laboratory environment. The second case consists of Case 1 with the inclusion of measurement error as a source of variation. The final case includes variation that results from selecting sites according to a probabilistic sampling scheme.

B. Sample Size: SA: Controlled Experiment

This first case (controlled experiment) is an obvious oversimplification of the actual measurement process. The composition of σ is kept as simple as possible. We assume that the sites included have, in essence, been constructed in a laboratory for the validation study. We assume that algorithm scoring is not subject to error. Under these assumptions, the validation study is similar to a controlled experiment conducted in a laboratory setting. The experimental units are sites which have been selected to reflect the desired distribution of SA scores.

Although this formulation is an oversimplification of the actual problem, it serves as a reference case for more realistic formulations. The fact that the complexities of a realistic model of variability are not considered allows us to focus clearly on the sample size issues.

1. SA Formulation 1

In Formulation 1, we consider two statistical hypotheses:

(a) $H_0: \mu_{10} = \mu_B$

$H_1: \mu_{10} > \mu_B$

and

(b) $H_0: \mu_{40} = \mu_{10}$

$H_1: \mu_{40} > \mu_{10}$.

Recall that SA scores of 10 and 40 have been designated as decision points for corrective action. Hypothesis (a) distinguishes airborne asbestos levels at sites where SA is 10 from background asbestos levels. Hypothesis (b) distinguishes between asbestos levels at sites where the SA scores are 10 and 40, respectively. Note that no specific airborne asbestos concentration levels are mentioned in the hypotheses. If specific levels were to be used, they would come from the dose-response curve that relates "risk" to asbestos dose level. The proposed validation tests (hypotheses (a) and

(b)) are only indirectly related to "risk". It is possible however that μ_{40} is found to be larger than μ_{10}, but both values are well below the acceptable level based on dose-response studies. On the other hand, H_0: $\mu_{40} = \mu_{10}$ may be accepted and both values may be large enough to pose a serious health risk. As mentioned earlier, the validation tests described in this report deal only with the relationship between air levels and exposure assessment algorithm scores. The other validation step linking air levels to health risk and establishing maximum safe dose levels is treated elsewhere. (See Support Document, Asbestos-Containing Materials in Schools, Health Effects and Magnitude of Exposure, USEPA, October 1980).

Hypothesis (a). We consider hypothesis (a) first. The statistical test of null hypothesis, H_0, will be carried out using the two sample "t" statistic,

$$t = (n/2)^{1/2} \, (\bar{Y}_{10} - \bar{Y}_B) \, / \, s \, ,$$

with 2n-2 degrees of freedom, and s is an estimate of σ using the pooled set of data. It is assumed that the number of sites of each type (S = 10 and background) is n. In order to calculate the required sample size, the following quantities must be specified:

(1) Significance level (α)--probability of choosing H_1 when H_0 is true;

(2) Specific alternative hypothesis--magnitude of the difference between μ_{10} and μ_B that is considered to be important; and

(3) power ($1 - \beta$)--probability of choosing a specific alternative, H_1, when it is correct.

In the planning stages of the experiment, it is often difficult to arrive at one particular set of the three values to be used in determining the sample size. Because of this difficulty, it is instructive to select a range of values and generate a table of corresponding sample sizes.

In this way, is it possible to develop a perspective on how significance level, alternative, power, and sample size are related.

Table 3 has been constructed for this purpose. The columns correspond to various values of $1 - \beta$, the power of the test. The rows (taken in pairs), correspond to various alternatives. The alternatives are differences between μ_{10} and μ_B stated in units of σ for convenience. For example, Rows 5 and 6 correspond to the alternative, H_1: $\mu_{10} - \mu_B = 2\sigma$; that is, the difference is two standard deviation units. The first row in each pair corresponds to a significance level of $\alpha = .05$. The second row in each pair is for $\alpha = .01$.

Note in general that larger sample sizes are required as (1) the power is increased, and (2) K is decreased (i.e., the difference between μ_{10} and μ_B is decreased). In Row 6 we see that with $\alpha = .01$, it takes 10 sites with SA scores of 10 and 10 background sites to have a 90 percent chance of finding a difference between μ_{10} and μ_B of two standard deviation units ($\mu_{10} - \mu_B = 2\sigma$). Increasing the power from 90 to 99 percent requires 5 additional sites of each type. However, holding the power at 90 percent and changing the alternative to a difference of one standard deviation unit (Row 4) requires an increase in the sample size of 19 sites of each type.

If the value of σ were known, the determination of required sample size would be simplified by using the known value of σ to state the alternative in terms of the original measurement units, ng/m^3 in this case. If the value of σ is not known, it is instructive to find other representations that can be used to help interpret the alternative hypotheses. Many representations are possible. The one selected for discussion here is based on a probability. Consider two measurements Y_{10} and Y_B from one site where the SA score is 10 and a background site, respectively. It is expected that $Y_{10} \geq Y_B$. Based on our assumptions concerning the structure

Table 3. Required Sample Size (n) For Testing H_0: $\mu_{10} = \mu_B$ Versus H_1: $\mu_{10} - \mu_B = K\sigma$ (n sites in each group)

Standard Deviation Units (K)	Significance Level (a)	Power (1 - β)				
		.50	.80	.90	.95	.99
0.5	$\alpha = .05$	23	57	70	90	>100
	$\alpha = .01$	45	80	>100	>100	>100
1.0	$\alpha = .05$	7	15	19	23	35
	$\alpha = .01$	15	23	29	34	49
2.0	$\alpha = .05$	3	5	6	7	10
	$\alpha = .01$	6	8	10	12	15
3.0	$\alpha = .05$	2	4	4	5	6
	$\alpha = .01$	4	6	6	7	8
4.0	$\alpha = .05$	2	2	3	4	4
	$\alpha = .01$	4	5	5	6	6

Source: Handbook of Statistical Tables, D.W. Owen, Addison-Wesley, 1962, Section 2.2.

of the data, the probability of the "unexpected", namely $Y_{10} \leq Y_B$, depends only on K when $\mu_{10} - \mu_B = K\sigma$. This probability may be interpreted as a measure of "overlap" between the two distributions of airborne asbestos concentration levels. Table 4 shows values of the probability that $Y_{10} \leq Y_B$ corresponding to different values of K.

Alternative hypotheses of the form H_1: $\mu_{10} - \mu_B = K\sigma$ are interpreted in terms of the "overlap" measure as follows. When K = 2, the overlap is small: .079. A minimal amount of data should be required to distinguish between the two means when the overlap of the two distributions are so small. When K = .5, the overlap is substantial. In this case, a large amount of data would be required to distinguish between the means when the overlap of the two distributions is that large. For the numerical example given above, when α = .01, it takes 10 sites of each type to have a 90 percent chance of finding a difference between μ_{10} and μ_B when the overlap is .079. However, for the same significance level and power, it takes 29 sites of each type to find a difference between μ_{10} and μ_B when the overlap is equal to .239.

The analysis of sample size requirements in terms of multiples of σ is instructive, and introducing the "overlap" measure is helpful, but it would be more informative to compare these requirements relative to the actual units of measurement--ng/m^3. In order to accomplish that goal, it is necessary to have preliminary estimates of σ and μ_B. Data that are presented in the report "Measurement of Asbestos Air Pollution Inside Buildings Sprayed With Asbestos" (Sebastien 1980) have been used to obtain estimates. In that report, a distribution of background values is given for Paris. The report indicates that the lognormal distribution is appropriate for background values. An estimate of μ_B = -.755 is given, and it is stated that 99 percent of the values are less than or equal to 7 ng/m^3. Solving the equation

$$\mu_B + 2.33\,\sigma = \ln 7$$

Table 4. Probability That $Y_{10} \leq Y_B$ When Testing $H_0: \mu_{10} = \mu_B$ Versus $H_1: \mu_{10} - \mu_B = K\sigma$

Standard Deviation Units (K)	$P(Y_{10} \leq Y_B)$ [a]
.5	.363
1	.239
2	.079
3	.017
4	.002

a Based on Assumptions stated in Section IV.B, $P(Y_{10} < Y_B) = \Phi(-K/\sqrt{2})$, where $\Phi(\cdot)$ is the Standard Normal Distribution Function.

yields an estimate of σ equal to 1.159. Table 5 shows the numerical values of the alternatives in ng/m^3 corresponding to values of the multiple, K.

Using the example introduced above, if $\alpha = .01$, it takes measurements on 10 sites of each type to have a 90 percent chance of detecting a difference of 8.4 ng/m^3. For the same significance level and power, it takes 29 sites of each type to correctly identify a difference of 2 ng/m^3 (refer to Tables 3 and 5). This example which is based on an estimate of σ from the Paris data (Sebastien 1980) suggests that the statistical test of H_0: $\mu_{10} = \mu_B$ is very sensitive to small differences between μ_{10} and μ_B. Requiring a total of only 20 sites to detect a difference of 8.4 ng/m^3 seems remarkable. The difficulty, if there is any difficulty, may be that the estimate of σ is too small. The estimate may be too small either because the sample on which it is based is not truly representative of background concentration levels, or because the assumption that the value of μ is independent of σ is incorrect. One plausible alternative assumption is that there is a value, σ_B, for background measurements and another value, σ, for all other concentration levels. This assumption adds an element of complication to the statistical test used for the hypothesis H_0: $\mu_{10} = \mu_B$; however, tests for the other hypotheses are unaffected. It is to be expected that σ is larger than σ_B. Therefore, in our sample size determination planning effort, it is conservative to consider values of σ larger than 1.159 (obtained from the Sebastien 1980 data).

Table 6 shows examples of how the numerical alternatives change when σ is increased to 2. Referring once again to the example introduced above, if $\alpha = .01$, it takes 10 sites of each type to have a 90 percent chance of detecting a difference of 186.1 ng/m^3 between μ_{10} and μ_B. For the same significance level and power, 29 sites of each type are re-

Table 5. Raw Measurement Units (ng/m^3) For The Alternative Hypothesis When Testing H_0: $\mu_{10} = \mu_B$ Versus H_1: $\mu_{10} - \mu_B = K\sigma$

Standard Deviation Units (K)	Raw Measurement Units [a] ng/m^3
.5	.7
1	2.0
2	8.4
3	28.9
4	94.0

a Obtained as $e^{\mu_B + \sigma^2/2}\left[e^{K\sigma} - 1\right]$, rounded to the nearest tenth. Values of $\mu_B = -.755$ and $\sigma = 1.159$ were estimated from Sebastien, 1980.

Table 6. Effect of Increased Standard Deviation On Raw Measurement Units (ng/m^3) For The Alternative Hypothesis When Testing $H_0: \mu_{10} = \mu_B$ Versus $H_1: \mu_{10} - \mu_B = K\sigma$

Standard Deviation Units (K)	Raw Measurement Units, ng/m^3 [a]	
	$\sigma = 1.159$	$\sigma = 2.0$
.5	.7	6.0
1	2.0	22.2
2	8.4	186.1
3	28.9	1397.6
4	94.0	10349.2

a Obtained as $e^{\mu B + \sigma^2/2}\left[e^{K\sigma} - 1\right]$, rounded to the nearest tenth.

quired to correctly identify a difference of 22.2 ng/m^3 (refer to Tables 3 and 6).

2. SA Formulation 1: Summary

It is clear from the preceding discussion that the process of arriving at a required sample size involves a number of decisions concerning parameter values. There is an element of subjectivity associated with each decision. The most difficult decision is choosing a fixed alternative. As was presented above, there are various ways to quantify an alternative. It may be equivalently specified as (1) a multiple of σ, (2) the probability of overlap ($P[Y_S - Y_B]$), or (3) in raw measurement units of nanograms per cubic meter, as well as other ways that have not been introduced here. Since there is usually no one quantification that is easily interpreted in all cases, it is useful to consider various specifications.

The question remains: "What sample size?" or more fundamentally, "How does one select sample size?" The task is best accomplished by first understanding the tradeoffs among the relevant parameters listed above. A summary table, Table 7, has been constructed for this purpose. The table displays power levels corresponding to a cross classification of proposed sample sizes and plausible alternatives. Four quantifications of each alternative are listed. If we settle on a significance level of 5 percent ($\alpha = .05$), it follows that with 16 or more sites of each type, most reasonable validation objectives will be met. Note that increasing n to either 24 or 32 is not sufficient to have a test that is sensitive to alternatives corresponding to K = .5. A sample size of n = 16 may be problematic if the true values of σ is larger than 2. If σ is as large as 2.5 (not included in the table), the alternative corresponding to K = 2 takes the value 1576.9 ng/m^3. In this case ($\sigma = 2.5$), a value of n approximately equal to 30 would be necessary to meet the objectives that are satisfied by n = 16 when $\sigma = 2$.

Table 7. Summary of Power (%) For Testing H_0: $\mu_{10} = \mu_B$ Versus H_1: $\mu_{10} - \mu_B = K\sigma$

n (sites per group)	K (std. dev. units)	Alternative Hypotheses [a]				
		.5	1	2	3	4
	$P(Y_B > Y_{10})$	.363	.239	.079	.017	.002
	ng/m^3, σ=1.159	.7	2.0	8.4	28.9	94.0
	ng/m^3, σ=2.0	6.0	22.2	186.1	1397.6	10349.2
4	$\alpha = .05$	<50	<50	65	90	95
	$\alpha = .01$	<50	<50	<50	50	50
8	$\alpha = .05$	<50	54	96	>99	>99
	$\alpha = .01$	<50	<50	80	99	>99
16	$\alpha = .05$	<50	83	>99	>99	>99
	$\alpha = .01$	<50	54	>99	>99	>99
24	$\alpha = .05$	50	95	>99	>99	>99
	$\alpha = .01$	<50	82	>99	>99	>99
32	$\alpha = .05$	60	98	>99	>99	>99
	$\alpha = .01$	<50	93	>99	>99	>99

a Linear Interpolation From Table 3.

It is clearly essential to have a reliable preliminary estimate of σ in order to arrive at a judicious choice of n.

Hypothesis (b). Hypothesis (b) states that H_0: $\mu_{40} = \mu_{10}$ is to be tested against H_1: $\mu_{40} > \mu_{10}$. The test statistic is similar to the statistic used in Hypothesis (a), namely,

$$t = (n/2)^{1/2} \, (\bar{Y}_{40} - \bar{Y}_{10}) \, / \, s \, ,$$

which has a "t" distribution with 2n-2 degrees of freedom. The characteristics of the test are found in Tables 3 and 7. In the current formulation of the validation problem, the SA designations of "background," S = 10 and S = 40 serve only as labels. Therefore, Hypotheses (a) and (b) have identical characteristics. The discussion presented for Hypothesis (a) applies without modification to Hypothesis (b).

3. SA Formulation 2

We assume that the relationship between Z, $(Z = Y_S - Y_B)$, and the SA score (denoted by S) can be approximated as a linear function

$$Z = \beta_0 + \beta_1 \cdot S + \varepsilon$$

where ε is a random quantity having an expected value equal to zero and variance equal to σ^2. Since we continue to treat the data as though they were generated in a controlled laboratory setting, σ^2 is a composite of variation in Z and variation associated with errors due to linear approximation. S is assumed to be measured without error.

The hypothesis of interest is

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 > 0 \, .$$

In this formulation, the acceptance of H_1 means that the relationship has been validated for all values of the SA score. As with Formulation 1, the alternative hypothesis must be made specific in order to arrive at the required sample size. The alternative considered is similar to those

formulated earlier. In particular, we shall look at hypotheses such as H_0: $\mu_{40} = \mu_{10}$. Note that the model implies that

$$\mu_S - \mu_B = \beta_0 + \beta_1 \cdot S \quad .$$

Therefore, the hypothesis that $\mu_{40} = \mu_{10}$ is equivalent to $30\beta_1 = 0$ or $\beta_1 = 0$. The alternative, H_1: $\mu_{40} - \mu_{10} = K\sigma$ is equivalent to H_1: $30\beta_1 = K\sigma$ or H_1: $\beta_1 = K\sigma/30$. It follows that the hypotheses considered in the previous section correspond to special cases of the more general hypothesis, H_0: $\beta_1 = 0$. However, it is demonstrated below that the characteristics--sample size and power--of the statistical test of H_0: $\beta_1 = 0$ are different from the characteristics of the statistical test used in Formulation 1. As expected, there is a reduction in sample size resulting from the assumption that the relationship between Z and SA is linear.

Test Characteristics: H_0: $\beta_1 = 0$

The test of H_0 is based on a "t" statistic defined as

$$t = \frac{b_1}{(s/\Sigma_S)}$$

with n-2 degrees of freedom where:

(1) b_1 and s are standard regression model estimators of β_1 and σ, respectively, and

(2) $$\Sigma_S = \left[\sum_{i=1}^{n} (S_i - \bar{S})^2\right]^{1/2}, \{S_i\}$$

are SA scores at sites in the sample.

The quantity, Σ_S, plays a major role in determining sample size and power for the statistical test. For any fixed sample size, the power increases with increasing Σ_S.

We consider two data collection designs mentioned in the Battelle Study Design report (November 20, 1980) and a third design, the one that is optimal with respect to the assumed linear model. The Battelle report suggests collecting data (1) according to a 2^5 full-factorial design or (2)

according to a half-fraction of the 2^5 design. There are sound arguments for using one or the other of these designs that are not based on formal quantitative characteristics. Those arguments appear in the Battelle Study Design report and will be discussed later in this report. For the present, we analyze the characteristics of the design relative to the validation test of Formulation 2.

Design: 2^5 Full-Factorial and Half-Fraction. The description of sites that make up the 2^5 design is given in Table 8. The rows that are marked (*) are those that are designated for the half-fraction design. In order to analyze the characteristics of these designs, it is necessary to compute an SA score for each site and then compute the value of Σ_S. Although the design is based on factor scores that are dichotomous, the actual sites will be scored according to the original SA weights shown in Table 1. If the original weights were used to specify unique sites, there would be 324 unique combinations. By dichotomizing 5 of the factors (and ignoring others), each of the 324 sites is associated with one of the 32 sites shown in Table 8. Factor scores for each of those 32 types of sites have been calculated as averages of the weights that are combined by the dichotomization process. These average factor scores and their associated SA scores are presented in Table 9. A value of Σ_S has been computed for each design. For the full-factorial design, n = 32, Σ_S = 138.8; for a half-fraction design, n = 16, Σ_S = 98.2.

A table showing power of the statistical test of H_0: $\beta_1 = 0$ has been developed for each design. Tables 10 and 11 have the results for the full-factorial design and the half-fraction design, respectively. Recall that the alternative hypotheses are designated to detect differences in mean values of airborne concentration levels of asbestos corresponding to different values of SA. We considered H_1: $\mu_{40} - \mu_{10} = K\sigma$ which is equivalent to H_1: $\beta_1 = K\sigma/30$.

Table 8. Full-Factorial Sampling Plan

Site No.	Algorithm Factors[a]				
	Content	Friability	Condition	Exposure	Accessibility
1	Low	Low	Good	Low	Low
*2	Low	Low	Good	Low	High
*3	Low	Low	Good	High	Low
4	Low	Low	Good	High	High
*5	Low	Low	Bad	Low	Low
6	Low	Low	Bad	Low	High
7	Low	Low	Bad	High	Low
*8	Low	Low	Bad	High	High
*9	Low	High	Good	Low	Low
10	Low	High	Good	Low	High
11	Low	High	Good	High	Low
*12	Low	High	Good	High	High
13	Low	High	Bad	Low	Low
*14	Low	High	Bad	Low	High
*15	Low	High	Bad	High	Low
16	Low	High	Bad	High	High
*17	High	Low	Good	Low	Low
18	High	Low	Good	Low	High
19	High	Low	Good	High	Low
*20	High	Low	Good	High	High
21	High	Low	Bad	Low	Low
*22	High	Low	Bad	Low	High
*23	High	Low	Bad	High	Low
24	High	Low	Bad	High	High
25	High	High	Good	Low	Low
*26	High	High	Good	Low	High
*27	High	High	Good	High	Low
28	High	High	Good	High	High
*29	High	High	Bad	Low	Low
30	High	High	Bad	Low	High
31	High	High	Bad	High	Low
*32	High	High	Bad	High	High

a Content (high) $\geq$ 50%
Friability (high) > 1 (see Table 1 for codes)
Condition (Bad) = 5
Exposure (high) = 4
Accessibility (high) = 3

* denotes sites designated for the half-fraction design.

Table 9. Full-Factorial Sampling Plan, Average Factor Scores

Site No.	Algorithm Factors						
	Content	Friability	Condition	Exposure	Accessibility	Other[a]	SA Score[b]
1	2	1.0	1	.5	.5	2.5	9.0
*2	2	1.0	1	.5	3.0	2.5	14.0
*3	2	1.0	1	4.0	.5	2.5	16.0
4	2	1.0	1	4.0	3.0	2.5	21.0
*5	2	1.0	5	.5	.5	2.5	17.0
6	2	1.0	5	.5	3.0	2.5	22.0
7	2	1.0	5	4.0	.5	2.5	24.0
*8	2	1.0	5	4.0	3.0	2.5	29.0
*9	2	2.5	1	.5	.5	2.5	22.5
10	2	2.5	1	.5	3.0	2.5	35.0
11	2	2.5	1	4.0	.5	2.5	40.0
*12	2	2.5	1	4.0	3.0	2.5	52.5
13	2	2.5	5	.5	.5	2.5	42.5
*14	2	2.5	5	.5	3.0	2.5	55.0
*15	2	2.5	5	4.0	.5	2.5	60.0
16	2	2.5	5	4.0	3.0	2.5	72.5
*17	3	1.0	1	.5	.5	2.5	13.5
18	3	1.0	1	.5	3.0	2.5	21.0
19	3	1.0	1	4.0	.5	2.5	24.0
*20	3	1.0	1	4.0	3.0	2.5	31.5
21	3	1.0	5	.5	.5	2.5	25.5
*22	3	1.0	5	.5	3.0	2.5	33.0
*23	3	1.0	5	4.0	.5	2.5	36.0
24	3	1.0	5	4.0	3.0	2.5	43.5
25	3	2.5	1	.5	.5	2.5	33.75
*26	3	2.5	1	.5	3.0	2.5	52.5
*27	3	2.5	1	4.0	.5	2.5	60.0
28	3	2.5	1	4.0	3.0	2.5	78.75
*29	3	2.5	5	.5	.5	2.5	63.75
30	3	2.5	5	.5	3.0	2.5	82.5
31	3	2.5	5	4.0	.5	2.5	90.0
*32	3	2.5	5	4.0	3.0	2.5	108.75

a Average of air moving system and activity (see Table 1).

b SA score given in Table 1.

Table 10. Formulation 2: Full-Factorial Design
H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/\Delta$, n = 32,
Degrees of Freedom = 30, $\sum_S = 138.8$;
Body of Tables Gives Power of the Test (%)

K	Significance Level	Δ				
		5[a]	10	20	30	40
.5	α = .05	**	**	97	75	52
	α = .01	**	**	85	45	25
1	α = .05	**	**	**	99	97
	α = .01	**	**	**	98	85
2	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	**
3	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	**
4	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	**

** Power exceeds 99%.

a Recall Δ = 5 implies H_0: $\mu_{s+5} = \mu_s$;

H_1: $\mu_{s+5} - \mu_s = K\sigma$.

Source: Handbook of Statistical Tables, D.W. Owen, Addison-Wesley, 1962, Section 2.2.

Table 11. Formulation 2: Half-Fraction Design
H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/\Delta$, n = 16,
Degrees of Freedom = 14, $\sum_S = 98.2$;
Body of Table Gives Power of the Test (%)

K	Significance Level	Δ				
		5[a]	10	20	30	40
.5	α = .05	**	**	75	45	25
	α = .01	**	99	45	20	13
1	α = .05	**	**	**	92	75
	α = .01	**	**	99	72	45
2	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	99
3	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	**
4	α = .05	**	**	**	**	**
	α = .01	**	**	**	**	**

** Power exceeds 99%.

a Recall Δ = 5 implies H_0: $\mu_{s+5} = \mu_s$;

H_1: $\mu_{s+5} - \mu_s = K\sigma$.

Source: Handbook of Statistical Tables, D.W. Owen, Addison-Wesley, 1962, Section 2.2.

A more general statement of the testing problem is

$$H_0: \mu_{S+\Delta} = \mu_S$$

$$H_1: \mu_{S+\Delta} - \mu_S = K\sigma \ ,$$

which is equivalent to

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 = K\sigma/\Delta$$

in Formulation 2. Tables 10 and 11 show powers corresponding to values of K = .5, 1, 2, 3, 4, and Δ = 5, 10, 20, 30, 40.

Design: Optimal. If the assumption that the relationship between Z and SA is linear is accepted without question, then the most efficient use of resources is to choose sites in a way that maximizes $\sum_S$. $\sum_S$ is maximized by choosing one-half of the sites at the low end of the SA scale and the other half at the high end of the SA scale. The resulting value of $\sum_S$ is $49.9\ n^{1/2}$. Table 12 shows the power of the statistical test corresponding to sample sizes n = 4, 8, 16, and 32 where n/2 sites are at each end of the SA scale. The alternatives considered are $H_1: \beta_1 = K\sigma/\Delta$. The value of Δ has been fixed at 30 in Table 12.

The statistical test based on this design is more powerful than the tests based on either the full-factorial or the half-fraction designs. By comparing Tables 10, 11, and 12, we see that for any fixed alternative (K, Δ = 30), the power values read from Table 12 are at least as large as comparable values found in Tables 10 and 11. Equivalently, it is possible to meet power objectives with fewer sites if the optimal design is used. For example, if the alternative hypothesis is $H_1: \mu_{40} - \mu_{10} = .5\sigma$, then 32 sites are required to reach a power level of 75 percent with the full-factorial design (Table 10, Row 1). When using the optimal design, approximately 10 sites would be required for power to be equal to 75 percent (Table 12, Row 1).

Table 12. Formulation 2: Optimal Design
H_0: $\beta_1 = 0$; H_1: $\beta_1 = K\sigma/30$ (i.e., $\Delta = 30$),
$\sum_S = 49.9 \times n^{1/2}$ Body of the Table Gives
Power of the Test (%)

K	Significance Level	Sample Size (n)			
		4	8	16	32
.5	$\alpha = .05$	30	65	94	99
	$\alpha = .01$	7	35	75	99
1	$\alpha = .05$	30	65	94	99
	$\alpha = .01$	7	35	75	99
2	$\alpha = .05$	**	**	**	**
	$\alpha = .01$	**	**	**	**
3	$\alpha = .05$	**	**	**	**
	$\alpha = .01$	**	**	**	**
4	$\alpha = .05$	**	**	**	**
	$\alpha = .01$	**	**	**	**

** Power exceeds 99%.

The advantage enjoyed by the optimal design is not without consequence. In effect, an assumption--the linearity of the relationship between Z and SA scores--replaces the need for additional sites. The savings in sites is satisfying provided the assumption of linearity is correct. However, if the optimal design is used, no data will be obtained that are useful for assessing the assumption.

The types of systematic errors that may occur as a result of an incorrect assumption about linearity are depicted in Figure 4. Consider for example, the effects of an incorrect linearity assumption on the test of $\mu_{40} = \mu_{10}$. In Figure 4a, the true relationship is convex and the assumed relationship is linear. The change in airborne concentration levels corresponding to a change from 10 to 40 on the SA scale may be understated if the linear representation were accepted. If the true relationship is concave (Figure 4b), the change in airborne levels may be overstated. In both situations, the discrepancies could cause the outcome of the validation analysis to be in error.

4. Comparison: Formulation 1 Versus Formulation 2

Table 13 has been prepared to compare the efficacy of the different designs proposed under Formulation 1 and Formulation 2. A formal analysis of statistical power indicates that Formulation 2 is preferred and that the optimal design provides for the most efficient use of resources. However, there are other issues besides statistical power that must be considered in selecting a design.

As mentioned earlier, Formulation 2 is superior provided that the assumption of linearity between Z and SA is correct. But linearity is only an assumption. It is prudent to implement a design that is efficient (high level of statistical power) and that also incorporates information that makes it possible to check assumptions. Clearly the "optimal" design of Formulation 2 does not meet this last criterion. The optimality of the "optimal" design rests totally

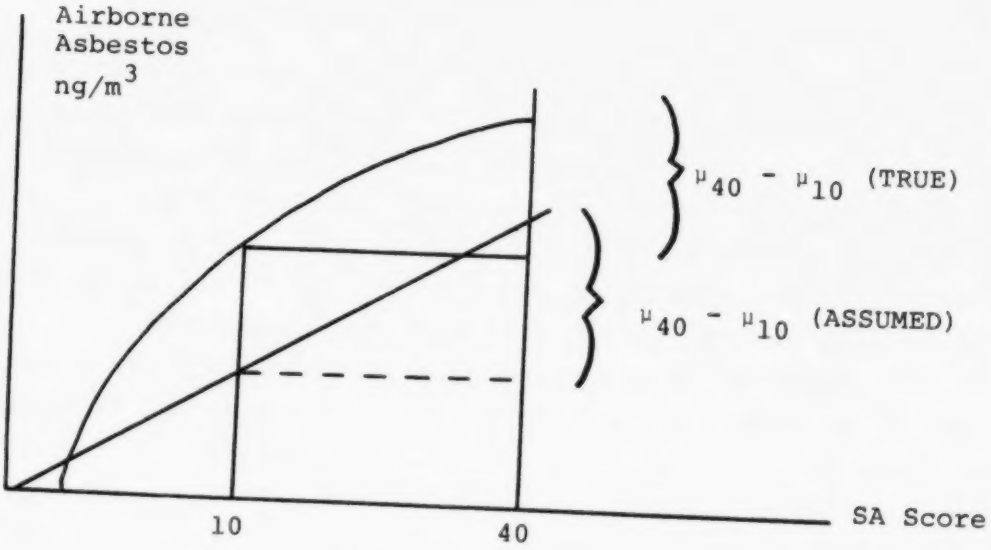


Figure 4a

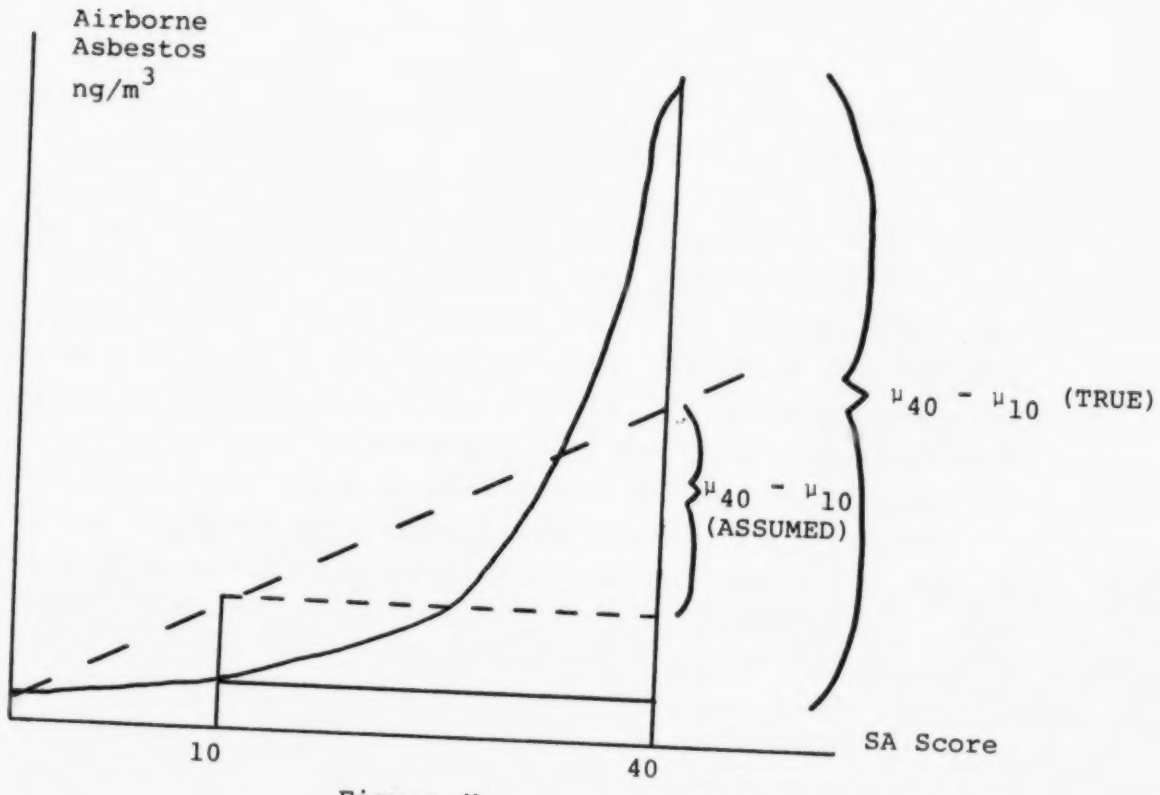


Figure 4b

Figure 4. Some effects of an erroneous linear assumption.

Table 13. Comparison: Formulation 1[a] Versus Formulation 2.
$\alpha = .05$, $\Delta = 30$; Body of Table Presents Power (%)

K	Formulation 1		Formulation 2			
			Full Factorial	Half Fraction	Optimal	
	n=32	n=16	n = 32	n = 16	n = 32	n = 16
.5	50	50	75	45	99	94
1	83	54	99	92	**	**
2	99	96	**	**	**	**
3	**	**	**	**	**	**
4	**	**	**	**	**	**

** Power exceeds 99%.

a Formulation 1 = H_0: $\mu_{s+\Delta} = \mu_s$ versus H_1: $\mu_{s+\Delta} - \mu_s = K\sigma$

Formulation 2 = H_0: $\beta_1 = 0$ versus H_1: $\beta_1 = K\sigma/30$

on the assumption of linearity. The full-factorial and half-fraction designs include data points that allow the linearity assumption to be checked. Each of these designs force the data to be judiciously scattered across the SA scale. The full-factorial design is most appealing. By covering all factor combinations, this design provides the best opportunity to observe any irregularities in the relationship between Z and SA.

It should be noted that the power of the proposed validity tests under Formulation 2 can be increased by deviating slightly from the formal full-factorial and half-fraction designs. Recall that power increases as Σ_S, the spread in the SA scores, increases. If the half-fraction design were chosen as a base, it could be supplemented with 16 additional sites having SA scores toward the extremes of the SA scale. That strategy would result in a design that retained the ability to assess departures from the linearity assumption and would simultaneously use the linearity assumption to provide a more powerful test than is available using the full-factorial design. A variety of adjustments to the basic designs (such as the one described above) are feasible. The rationale for adjusting the design is often a result of scientific judgment or practicality required during field implementation. In most cases, the adjusted designs can be evaluated using the techniques that led to Tables 10 through 13 so that it is possible to evaluate the information gain or loss associated with any adjustment.

C. Sample Size: Decision Tree Algorithm(DTA): Controlled Experiment

The validation test for DTA is based on the comparison of two groups of airborne asbestos concentration levels. The groups are defined in terms of an "action-deferred action" decision that results from applying the algorithm. DTA is based on four factors (see Figure 2) which determine 16 different types of sites. The classifi-

cation of sites into Group A (action) and Group DA (deferred action) is shown in Table 2. Five sites (8, 12, 15, 23, 32) are in Group A; eleven sites are in Group DA (2, 3, 5, 9, 14, 17, 20, 22, 26, 27, 29). The validation test has been formulated as

$$H_0: \mu_A = \mu_{DA}$$

versus

$$H_1: \mu_A > \mu_{DA}$$

where

$$\mu_A = \sum_{\text{Group A}} \mu_i \; / \; 5$$

$$\mu_{DA} = \sum_{\text{Group DA}} \mu_i \; / \; 11 \; .$$

Table 14 summarizes the power to sample size relationship associated with the alternative hypothesis formulated as $H_1: \mu_A = \mu_{DA} + K\sigma$. The first column of the table (labeled n = 16, df = 14) corresponds to the basic half-fraction design which allows for one site of each type in the sample. The remaining three columns describe the effects of increasing the sample size by replicating the design once, twice, and three times, respectively.

The replicated designs are preferred not only because they provide additional power, but because they provide more basic information than the unreplicated design (Column 1). If the basic 16 sites are sampled once, it is necessary to assume that the means for sites in Group A are identical and equal to μ_A. It is also necessary to assume that the means for sites belonging to Group DA are identical and equal to μ_{DA}. In addition, the variances for all sites are assumed to be equal. In this case, the statistical test is based on

$$t = \frac{\bar{Y}_A - \bar{Y}_{DA}}{s\left(\frac{1}{5} + \frac{1}{11}\right)^{1/2}}$$

with 14 degrees of freedom where

Table 14. Statistical Test: DTA; $H_0: \mu_A = \mu_{DA}$; $H_1: \mu_A - \mu_{DA} = K\sigma$
Body of Table Presents Power of Test (%)

K	Significance Level	Number of Replications (m)[a]			
		0 n=16 DF=14[b]	1 n=32 DF=16	2 n=48 DF=32	3 n=64 DF=48
.5	$\alpha = .05$	20	35	50	60
	$\alpha = .01$	10	15	25	30
1	$\alpha = .05$	55	80	95	98
	$\alpha = .01$	25	50	80	90
2	$\alpha = .05$	95	**	**	**
	$\alpha = .01$	85	99	**	**
3	$\alpha = .05$	**	**	**	**
	$\alpha = .01$	**	**	**	**
4	$\alpha = .05$	**	**	**	**
	$\alpha = .01$	**	**	**	**

a Number of times the basic 16 sites design is replicated.

b DF = degrees of freedom.

** Power exceeds 99%.

$$\bar{Y}_A = \sum_{\text{Group A}} Y_i / 5, \qquad \bar{Y}_{DA} = \sum_{\text{Group DA}} Y_i / 11$$

and

$$s^2 = \left[\sum_{\text{Group A}} (Y_i - \bar{Y}_A)^2 + \sum_{\text{Group DA}} (Y_i - \bar{Y}_{DA})^2 \right] \Big/ 14$$

If the design is replicated, there is sufficient data to estimate the mean for each type of site and to check on the validity of the assumption that all variances are equal. In this case, the test becomes

$$t = \frac{\sqrt{m}\left(\bar{Y}_A - \bar{Y}_{DA}\right)}{s\left(\frac{1}{5} + \frac{1}{11}\right)^{1/2}}$$

with 16 m degrees of freedom where m is the number of replications and where

$$\bar{Y}_A = \sum_{\text{Group A}} \sum_{\text{Site}} Y_{ij} / 5\ (m + 1)$$

$$\bar{Y}_{DA} = \sum_{\text{Group DA}} \sum_{\text{Site}} Y_{ij} / 11\ (m + 1)$$

and

$$s^2 = \left[\sum_{\text{Group A}} \sum_{\text{Site}} (Y_{ij} - \bar{Y}_i)^2 + \sum_{\text{Group DA}} \sum_{\text{Site}} (Y_{ij} - \bar{Y}_i)^2 \right] \Big/ 16m$$

with $\bar{Y}_i$ being the site mean. Therefore, the argument for taking at least one replication is not based totally on power considerations. At least one replication is required to keep from having to impose unrealistic assumptions on the validation test. Considering statistical power, it appears from Table 14 that at least one replication is also required if the statistical test is to distinguish group differences as small as one standard deviation (K = 1).

D. Sample Size: Measurement Error

In the previous discussions of sample size and precision, it was assumed that sites are scored without error. To score without error means that a group of raters presented with a given site will produce identical ratings on all factors. This assumption was made in the previous

section to simplify the formulation of the statistical problem so that the presentation of sample size issues would not be clouded. However, the assumption that exposure scores are produced without error is not realistic. Previous work (Battell, 1980; Price et al. 1980) strongly suggests that different raters score a given site differently. In this section we formulate the validation problem in a way that allows for variation due to error in exposure assessment scoring. For validation of SA, Formulation 2, the linear model approach is reanalyzed allowing for scoring errors. For validation of DTA, a model using composite distributions is presented and analyzed. In both cases, the effect of errors in assessment exposure scoring reduces the power of the validation test associated with any fixed sample size.

1. SA: Formulation 2

We assume that the relationship between airborne asbestos concentration levels and the algorithm score is linear. The formulation of the model that is used in the validation test when errors are present in scoring is slightly different than the model when scoring is without error. The model formulation follows.

Assume that the observed values of Z and SA are given by

$$z_i = Z_i + \varepsilon_i, \quad s_i = S_i + \eta_i$$

where ε and η are random quantities, uncorrelated, with expected values zero and variances σ_ε^2 and σ_η^2, respectively. The relationship between Z and SA is specified as

$$Z = \beta_0 + \beta_1 \cdot S + \varepsilon \ .$$

The validation test is formulated as before, H_0: $\beta_1 = 0$ versus H_1: $\beta_1 + K\sigma/\Delta$, where $\sigma = \sigma_\varepsilon$. The test is based on an estimate of β_1 and the standard error associated with that estimate. The formulation given above follows Britt and Luecke 1973. An estimate, $\hat{\beta}_1$ of β_1 is obtained by constrained maximum likelihood estimation with

$$\text{Var } (\hat{\beta}_1) = \frac{(R\ \beta_1^2 + 1)\ \sigma_\varepsilon^2}{\sum_S^2}$$

where $\sum_S^2 = \sum(S_i - \bar{S})^2$ and $R = \sigma_\eta^2/\sigma_\varepsilon^2$. (Note that when R=0, $\sigma_\eta^2 = 0$ which means that SA is measured without error; then Var $(\hat{\beta}_1) = \sigma_\varepsilon^2 / \sum(S_i - \bar{S})^2$ which is the variance of $\hat{\beta}_1$ in the classical regression model.)

In the case when $R \neq 0$, the variance of $\hat{\beta}_1$ under the alternative hypothesis that $\beta_1 \neq 0$ depends on the actual value of β_1 being considered. The dependence of the variance on the true value of the parameter in question makes the evaluation of the power of the statistical test difficult. The variance of $\hat{\beta}_1$ when there is error in the algorithm score, is larger than the variance of $\hat{\beta}_1$ when there is no error. When arguing qualitatively, because of the increase in variance, it is expected that more sites would be required to maintain the power of the validation test when errors are present than when algorithm scoring is error free. We have chosen not to assess the required increase in sites quantitatively.* The necessary calculations are outlined in Appendix A. In the next section, the effects of errors in algorithm scoring are analyzed with respect to validation of DTA. Tables showing the degradation of statistical power are included for that analysis.

2. DTA

When attempting to validate DTA, errors in scoring the algorithm factors may lead to the misclassification of a site into the action group when deferred action is correct, or conversely into the deferred action group when action is the correct decision. If the probability

---

* This analysis requires direct calculation of probabilities associated with the "noncentral t" distribution. All other analyses of sample size have been based on published tables of the noncentral t. It is not within the scope of this project to develop computer programs to make these calculations.

that misclassification occurs approaches one-half, it becomes impossible to validate DTA. In order to demonstrate the effects of errors in algorithm scoring (equivalently, misclassification of sites), we formulate the validation problem as a hypothesis testing problem involving mixtures of distributions. This formulation is presented below followed by an analysis of the degradation in statistical power associated with misclassification of sites.

Formulation. We first introduce the notation that is necessary to describe the parameters that must be considered in the validation analysis. Let P(A/DA) be the probability of designating a site for Group A (action) when it should be in Group DA, and let P(DA/A) be the probability of designating a site for Group DA when it correctly belongs to Group A. For convenience, it is assumed that these probabilities are equal:

$$p = P(A/DA) = P(DA/A) \ .$$

Let μ_A and σ^2 (μ_{DA} and σ^2) denote the mean and variance of the sites that belong to Group A (Group DA). These symbols designate the parameter values when there is no misclassification, $p = 0$. If $p \neq 0$, ther for Group A,

$$\text{Mean:} \quad \eta_A = (1 - p)\ \mu_A + p\ \mu_{DA}$$

$$\text{Variance:} \quad \sigma_A^2 = \sigma^2 + p(1 - p)\ (\mu_A - \mu_{DA})^2$$

and for Group DA,

$$\text{Mean:} \quad \eta_{DA} = p\ \mu_A + (1 - p)\ \mu_{DA}$$

$$\text{Variance:} \quad \sigma_{DA}^2 = \sigma^2 + p(1 - p)\ (\mu_A - \mu_{DA})^2 \ .$$

As described earlier in this report, validation is equivalent to testing H_0: $\mu_A = \mu_{DA}$ against H_1: $\mu_A - \mu_{DA} = K\sigma$. However, because of the possibility of misclassification, the actual statistical test is

$$H_0: \ \eta_A = \eta_{DA}$$

versus

$$H_1: \ \eta_A - \eta_{DA} = K\sigma \ .$$

Note that H_0: $\eta_A = \eta_{DA}$ is equivalent to H_0: $\mu_A = \mu_{DA}$. For the alternative, our interest is in H_1: $\mu_A - \mu_{DA} = K\sigma$ which is equivalent to

$$H_1: \eta_A - \eta_{DA} = (1 - 2p)\, K\sigma \ .$$

This last statement of the alternative partially shows the effect of misclassification on the validation test. The alternative of interest, $K\sigma$, is reduced by the factor (1 - 2p) because we are forced to deal with distributions that have been contaminated by misclassification of sites. We know that for any fixed power level, reduction of the magnitude of the alternative leads to the necessity to increase the sample size (number of sites). As the probability of misclassification increases, the requirement for additional sites to meet a stated validation objective intensifies. In the extreme case, $p = 1/2$, the probability of detecting any difference remains at the significance level independent of the sample size.

Degradation in Power of the Test. Table 15 has been prepared to indicate the effect that scoring errors have on the validation test. The format of Table 15 is similar to the format used in Table 14. Table 15 displays power levels corresponding to various assumptions about the probability of misclassification.* Table 14 contains only the error-free case, $p = 0$. These values from Table 14 are reproduced as the first row in each set of four rows in Table 15.

From Table 15, we see that for a fixed sample size and a fixed alternative hypothesis (value of K), the power declines as the probability of misclassification, p, increas-

---

* As in the error-free formulation of the validation problem for the DTA, the power calculations are based on the "t" test for equality of means. Since the distributions under consideration are mixtures, the statistical test is an approximation. However, it is believed to be sufficiently accurate to demonstrate the effect that misclassification has on power levels.

Table 15. DTA: $H_0: \eta_A = \eta_{DA}$; $H_1: \eta_A - \eta_{DA} = (1-2p)K\sigma$

Body of the Table Gives Power of Degradation When Sites Are Misclassified. Significance Level = .05

K	Probability of Misclassification (p)	Number of Replications (m)			
		0 n = 16 DF = 14	1 n = 32 DF = 16	2 n = 48 DF = 32	3 n = 64 DF = 48
.5	0	20	35	50	60
	.1	20	25	40	40
	.25	10	10	20	25
	.4	5	5	10	10
1	0	55	80	95	98
	.1	40	65	80	90
	.25	20	25	45	55
	.4	10	10	15	15
2	0	95	**	**	**
	.1	80	95	**	**
	.25	35	60	80	85
	.4	15	15	25	25
3	0	**	**	**	**
	.1	90	**	**	**
	.25	45	75	90	95
	.4	15	20	25	30
4	0	**	**	**	**
	.1	98	**	**	**
	.25	55	80	90	95
	.4	15	25	30	40

DF = degrees of freedom.
n = number of sites.
** Power exceeds 99%.

es. For example, when K = 1, 32 sites yield a power of 80 percent if algorithm scoring is done without error (p = 0). If p = .1, the power drops to 65 percent. It would require 48 sites to maintain the power of that test at approximately the 80 percent level when p = .1. Other examples may be extracted from the table. The important observation is that variability in algorithm scoring that leads to misclassification of sites can have a material impact on the strength of the validation results. For both the validation study and the ultimate utilization of the algorithm in the field, it is essential to implement training and techniques that will minimize scoring variation.

V. DESCRIPTION OF AVAILABLE SITES IN MONTGOMERY COUNTY AND NEW YORK CITY

In this section a description of algorithm scores in two school districts where the proposed field study could be carried out is presented. These two districts were selected because RTI had current data on Sawyer Algorithm scores for schools in these districts.

A. The Montgomery County Data

Two hundred schools were inspected by sanitarians from the Montgomery County, Ohio, Health District. Friable, asbestos-containing materials were identified in 75 sites in 23 schools. Table 16 presents the distribution of these original 75 sites in terms of a full-factorial sampling plan as described in Section IV (Table 8). The sites subject to some type of abatement activity since the original survey are identified in the footnotes. Inspection of Table 16 reveals that 12 sites have been subject to abatement activity. Currently, 33 percent (21/63) of the available sites have low friability and low percent asbestos. About 55 percent (35/63) of the available sites have high friability and low percent asbestos. Only 3 percent of available sites (2/63) have high percent asbestos and low friability, and only 8 percent of the available sites (5/63) have high friability and high percent asbestos. The available data thus suggest that approximately 90 percent (56/63) of the possible sites in Montgomery County have suspect materials with relatively small amounts of asbestos and that only 7 sites have asbestos levels of greater than 30 percent. This lack of high asbestos sites certainly argues against the initiation of a validation study in this area.

B. The New York City Data

The Division of School Buildings of the New York City Board of Education became aware of a potential asbestos exposure problem in January of 1977 shortly after six New Jersey elementary schools were closed. During that year,

Preceding page blank

Table 16. Distribution of Montgomery County Sites

Stratum	Dichotomized Algorithm Variables[g] % Asbestos	Friability	Condition	Exposure	Accessibility	Sites Before Repair	Total Sites Available After Repair
1	Low	Low	Good	Low	Low	0	
2	Low	Low	Good	Low	High	0	
3	Low	Low	Good	High	Low	18	
4	Low	Low	Good	High	High	3	21
5	Low	Low	Bad	Low	Low	0	
6	Low	Low	Bad	Low	High	0	
7	Low	Low	Bad	High	Low	0	
8	Low	Low	Bad	High	High	3[a]	
9	Low	High	Good	Low	Low	0	
10	Low	High	Good	Low	High	0	
11	Low	High	Good	High	Low	34[b]	
12	Low	High	Good	High	High	3[c]	35
13	Low	High	Bad	Low	Low	0	
14	Low	High	Bad	Low	High	0	
15	Low	High	Bad	High	Low	4[d]	
16	Low	High	Bad	High	High	0	
17	High	Low	Good	Low	Low	0	
18	High	Low	Good	Low	High	0	
19	High	Low	Good	High	Low	3[e]	
20	High	Low	Good	High	High	2[f]	2
21	High	Low	Bad	Low	Low	0	
22	High	Low	Bad	Low	High	0	
23	High	Low	Bad	High	Low	0	
24	High	Low	Bad	High	High	0	
25	High	High	Good	Low	Low	0	
26	High	High	Good	Low	High	0	
27	High	High	Good	High	Low	5	
28	High	High	Good	High	High	0	5
29	High	High	Bad	Low	Low	0	
30	High	High	Bad	Low	High	0	
31	High	High	Bad	High	Low	0	
32	High	High	Bad	High	High	0	

a 3 sites fixed
b 4 sites fixed
c 1 site fixed
d 1 site fixed
e 1 site fixed
f 2 sites fixed

g % Asbestos (high) more than 30%
(see table 1 for codes)
Friability (high) more than 1
Condition (bad) = 5
Exposure (high) = 4
Accessibility (high) = 3

the specifications of 321 buildings in New York City that were constructed between 1956 and 1976 were reviewed. The specification review identified approximately 185 buildings with asbestos-containing materials. By January 1981, 1411 buildings from approximately 1000 schools were physically inspected. Bulk samples confirmed the presence of asbestos materials in 257 buildings. Conflicting laboratory results were obtained in an additional 5 buildings.

RTI has received survey data that were collected in 266 schools. In the borough-specific RTI files, there are complete algorithm data from 2,265 sites in the 266 schools; 948 sites have incomplete data. The actual percentage of asbestos estimated from the bulk samples is available for approximately 10 percent of the 3,213 sites in the RTI files.

Sites within the 1411 school buildings that have asbestos materials, which have not been eliminated by abatement activities, constitute the major portion of the population from which validation study sites should be selected. Table 17 presents the distribution of the original, and of currently available, sites in terms of the full-factorial sampling plan. It should be noted that sites with less than 1 percent asbestos have not been deleted from Table 17, although it could be argued that these sites should not be considered eligible for the validation study. Inspection of Table 17 suggests that 52 percent (639/1226) of the available sites have relatively low percentages of asbestos and low friability, while 10 percent (117/1226) of the available sites have high percentages of asbestos and high friability. Thirty-seven (457/1226) percent of the available observations have low asbestos percentages and high friability, while only 1 percent (13/1226) of available sites have high asbestos percentages and low friability. Tables 16 and 17 clearly show that compared to Montgomery County, New York City has many more sites available for the proposed validation study.

Table 17. Distribution of New York Sites[a]

Stratum	Dichotomized Algorithm Variables[c]					Sites[a]		Total Presently Available
	% Asbestos	Friability	Condition	Exposure	Accessibility	Original	Current[b]	
1	Low	Low	Good	Low	Low	8	2	
2	Low	Low	Good	Low	High	1	1	
3	Low	Low	Good	High	Low	685	555	
4	Low	Low	Good	High	High	102	66	639
5	Low	Low	Bad	Low	Low	1	0	
6	Low	Low	Bad	Low	High	0	0	
7	Low	Low	Bad	High	Low	11	4	
8	Low	Low	Bad	High	High	27	11	
9	Low	High	Good	Low	Low	549	303	
10	Low	High	Good	Low	High	9	0	
11	Low	High	Good	High	Low	192	120	
12	Low	High	Good	High	High	21	14	457
13	Low	High	Bad	Low	Low	0	0	
14	Low	High	Bad	Low	High	0	0	
15	Low	High	Bad	High	Low	12	6	
16	Low	High	Bad	High	High	20	14	
17	High	Low	Good	Low	Low	0	0	
18	High	Low	Good	Low	High	0	0	
19	High	Low	Good	High	Low	11	11	
20	High	Low	Good	High	High	2	2	13
21	High	Low	Bad	Low	Low	0	0	
22	High	Low	Bad	Low	High	0	0	
23	High	Low	Bad	High	Low	0	0	
24	High	Low	Bad	High	High	0	0	
25	High	High	Good	Low	Low	520	89	
26	High	High	Good	Low	High	0	0	
27	High	High	Good	High	Low	75	16	
28	High	High	Good	High	High	4	1	117
29	High	High	Bad	Low	Low	0	0	
30	High	High	Bad	Low	High	0	0	
31	High	High	Bad	High	Low	9	6	
32	High	High	Bad	High	High	6	5	

a missing data, 948 sites, sites with less than 1% asbestos are included.

b after abatement activity.

c % Asbestos (high) 50% or more (see Table 1 for codes)
Friability (high) greater than 1
Condition (bad) = 5
Exposure (high) = 4
Accessibility (high) = 3

In the New York City schools, data collection "sites" were defined by the nature of the asbestos-containing material and the situation where the material was found. Soft friable material was usually found in boiler rooms, fan rooms, music rooms, and cafeterias. Sprayed-on fireproofing material was usually found on structural steel which might be exposed, or hidden behind a suspended ceiling.

Troweled-on acoustic plaster was usually found in corridors and auditoriums. The current recommendation to taking one bulk sample for every 5,000 square feet of homogeneous material was not followed strictly. Since a classroom is approximately 700 square feet in area, the one sample per 5,000 square feet recommendation necessitates one sample for every seven classrooms. The Division of School Buildings reports that acoustic plaster (bulk sample) results can vary from room to room and floor to floor even though the plaster's appearance is homogeneous. But this variation may be a function of laboratory error and not where the bulk sample was obtained.

VI. SAMPLE DESIGN

A. Introduction

This chapter presents the probability sample design proposed to study the relationships of the algorithms to airbrone asbestos levels in schools. The sample design will provide valid inferences for the study area discussed below. Note that a purposive selection of schools and sites within schools is not proposed because it would not permit statistically valid inferences to be made for the entire study area; the results would apply just to those sites where data were collected. Also, there would be little assurance that the results were not controlled or biased by researcher preconceptions or objectives.

It is deemed feasible at this time that the study include only a geographically restricted area, as opposed to being a national study. The inferential ability of the information generated by the study is concomitantly restricted. It is still thought important, however, that a sample design be employed to allow conclusions to be drawn at the level of the study area. The study area is in fact a real-world situation, rather than some purposively or convenient collection of classrooms that may well be atypical of any real setting.

B. Study Area

The proposed study area includes all eligible sites in New York City public schools. As discussed in Section V of this report, reasons for using New York City public schools are the availability of data for use in sample selection and the variety of asbestos-containing sites present. Eligible sites within a school include classrooms, hallways, cafeterias, kitchens, gymnasiums, locker rooms, libraries, and auditoriums. A protocol for partitioning hallways into sites remains to be developed. Storage rooms and offices are not included as eligible sites, as suggested

Preceding page blank

in Price et al. (1980b). This suggestion was based on the suspected lower activity levels in these rooms.

C. Overview of the Sample Design

A two-stage sample design with stratification imposed on each stage is proposed. First-stage sampling units are public schools in New York City. Stratification of the first-stage frame (i.e., list of all schools) is first provided by their classification into the following three strata: (1) asbestos-containing schools (according to prior investigation), (2) schools with unknown asbestos content, and (3) all remaining schools (those believed to have no asbestos, including those that have had asbestos problems corrected). Using data collected in a survey of New York City public schools for the presence of asbestos by the New York City Asbestos Task Force, the class of asbestos-containing schools will be further stratified according to asbestos content, friability, condition, exposure, and accessibility. This will result in 16 strata in all. At this time it is thought that the first-stage sample will probably consist of 16 schools. In this case, the first-stage sample of schools will be allocated equally among the first-stage strata. Schools will be selected from the first-stage strata with probability proportional to size measures based on school enrollment.

The second-stage frame will consist of all eligible sites in the first-stage sample of schools. By visual inspection, a trained rater will rate each of these sites as to friability, condition, exposure, and accessibility. All possible low/high combinations of these four factors will form the second-stage strata. The sample size will be allocated among the second-stage strata proportional to the stratum totals of numbers of sites, with the restriction that at least one site be selected from each nonempty second-stage stratum. Within second-stage strata, sites will be selected with equal probability and without replacement. To

facilitate variance estimation, two independent samples (of equal size) will be selected from the second-stage frame. Possible total sample sizes are 32 sites, 48 sites, or 64 sites.

At each selected site, bulk samples of friable material will be collected, according to the guidance given in Lucas et al. (1980). Air sampling will be conducted as described in Price et al. (1980b). Prior to air sampling, an exact protocol for placement of samplers within a site will be developed. This protocol should provide standardization in the procedures used across all sites in order to avoid biasing the results of the study. Additionally, at each selected site, both trained and untrained raters will independently score all algorithm factors.

The sample design outlined above is a statistically valid design that will give estimates for the study area that are free from selection bias. Every public school in the study area has a known positive probability of selection, and every eligible site in the study area has a known positive probability of selection. Additionally, the replication at the second-stage facilitates variance estimation. This design provides reasonable assurance that virtually all of the 32 factorial cells (see Section IV) will be filled, provided there exist such factor level combinations in the study area. Supplementary selection procedures can be used in the event that a cell believed to be nonempty is not in the sample as desired.

D. Construction and Stratification of the First-Stage Frame

The first-stage frame consists of all public schools in New York City. A previous survey of New York City public schools for the presence of asbestos is described in Section V of this report. Based on the results of this survey and records of corrective action taken since that time, there are currently 69 public shools in New York

City known to have at least one asbestos-containing site. (Asbestos is said to be present at a site if the average asbestos concentration at the site exceeds 1 percent.) The remaining schools fall into one of the following categories: (1) asbestos content unknown, (2) asbestos found present but subsequently removed or corrected, or (3) asbestos thought to be not present. Table 18 shows the distribution of schools among these categories.

The first-stage frame will first be stratified into the following three classes: (1) asbestos-containing schools, (2) schools with unknown asbestos content, and (3) all remaining schools. The class of asbestos-containing schools will be further stratified according to asbestos content, friability, condition, exposure, and accessibility.

Table 19 shows the classification of asbestos-containing schools with respect to all possible low/high combinations of the five factors listed above. A school is placed in a given category if any surveyed site in that school exhibited the specified combination of factor levels. This means that it is possible for one school (with more than one surveyed site) to be in more than one category. For asbestos-containing schools, it is proposed that strata be constructed from all possible low/high combinations of the five factors, with the exception that only one stratum will include all factor combinations involving high asbestos and low friability. (A site having both high asbestos content and low friability is known to be a fairly rare occurrence (see Table 17)). This yields 25 strata of asbestos-containing schools. The strata will be filled in ascending order of the number of schools (not yet placed in strata) exhibiting the corresponding factor level combinations, with the restriction that each school belong to only one stratum. That is, if a stratum has only one school, then that school will be in that stratum and no other. Then if a stratum has two schools, those two schools will be in that stratum and no other, etc.

Table 18. Asbestos Content of New York City Public Schools

Asbestos Content	Number of Schools
Asbestos Known Present	69[a]
Asbestos Content Unknown	48[a]
Asbestos Found Present but Later Removed or Corrected	89[a]
Asbestos Thought Not Present	792
TOTAL	988[b]

a Data from survey of New York City public schools by the New York Asbestos Task Force and records of corrective action.

b 1978 Curriculum Information Directory.

Table 19. Distribution of Asbestos-Containing Public Schools in New York City[a] With Respect to Asbestos Content, Friability, Condition, Exposure, and Accessibility

Asbestos Content	Friability	Condition	Exposure	Accessibility	Number of Schools
Low	Low	Good	Low	Low	0
Low	Low	Good	Low	High	0
Low	Low	Good	High	Low	26
Low	Low	Good	High	High	6
Low	Low	Bad	Low	Low	0
Low	Low	Bad	Low	High	0
Low	Low	Bad	High	Low	0
Low	Low	Bad	High	High	4
Low	High	Good	Low	Low	6
Low	High	Good	Low	High	0
Low	High	Good	High	Low	27
Low	High	Good	High	High	4
Low	High	Bad	Low	Low	0
Low	High	Bad	Low	High	0
Low	High	Bad	High	Low	4
Low	High	Bad	High	High	5
High	Low	Good	Low	Low	1
High	Low	Good	Low	High	0
High	Low	Good	High	Low	1
High	Low	Good	High	High	1
High	Low	Bad	Low	Low	0
High	Low	Bad	Low	High	0
High	Low	Bad	High	Low	0
High	Low	Bad	High	High	0
High	High	Good	Low	Low	5
High	High	Good	Low	High	0
High	High	Good	High	Low	7
High	High	Good	High	High	1
High	High	Bad	Low	Low	0
High	High	Bad	Low	High	0
High	High	Bad	High	Low	4
High	High	Bad	High	High	2
Missing					1

Data from survey of New York City public schools and records of corrective action.

A school was placed in one of the above categories if any surveyed site in that school exhibited the combination of factor levels corresponding to that category. It is possible for one school to be in more than one category.

a 69 schools in New York City are currently known to have greater than 1 percent asbestos.

Table 20 lists the 27 first-stage strata. The 25 strata of asbestos-containing schools are listed in the order in which they were filled, using the procedure discussed above. The number of schools in each stratum is given. Note that 11 strata contain no schools; of course no schools can be selected from these strata. The fact that these strata are empty does not mean, however, that no sites having any of these factor level combinations will appear in the sample. A site with one of these factor level combinations may exist as a nonsurveyed site in an asbestos-containing school, or such a site may exist in a school in stratum 15 or stratum 16. Any such site will have a known positive probability of appearing in the sample.

E. Allocation and Selection of the First-Stage Sample

At this time it is thought that the first-stage sample will probably consist of 16 schools. In this case, the first-stage sample of schools will be allocated equally among the 16 first-stage strata; i.e., one school will be selected from each of the first-stage strata.

It is proposed to select schools from the first-stage strata with probability proportional to size measures based on school enrollment. Information on the number of eligible sites per school is not readily available; however, it is thought that school enrollment will have a fairly strong positive relationship with the number of sites in the school. The idea is to have selection probabilities at the first-stage be such that a self-weighting sample can be obtained, to the extent possible given the design constraints. The distribution of New York City public schools by enrollment categories is shown in Table 21. Note in Table 20 that two strata, 1 and 11, contain only one school. Those two schools will be in the first-stage sample with certainty.

Table 20. First-Stage Strata

Stratum	Description					No. of Schools[a]
	Asbestos Content	Friability	Condition	Exposure	Accessibility	
	Low	Low	Good	Low	Low	0
	Low	Low	Good	Low	High	0
	Low	Low	Bad	Low	Low	0
	Low	Low	Bad	Low	High	0
	Low	Low	Bad	High	Low	0
	Low	High	Good	Low	High	0
	Low	High	Bad	Low	Low	0
	Low	High	Bad	Low	High	0
	High	High	Good	Low	High	0
	High	High	Bad	Low	Low	0
	High	High	Bad	Low	High	0
1	High	High	Good	High	High	1
2	High	Low	---	---	---	2
3	High	High	Bad	High	High	2
4	Low	Low	Bad	High	High	4
5	Low	Low	Good	High	High	4
6	Low	High	Good	High	High	3
7	Low	High	Bad	High	High	3
8	Low	High	Bad	High	Low	3
9	High	High	Bad	High	Low	4
10	High	High	Good	Low	Low	4
11	High	High	Good	High	Low	1
12	Low	High	Good	Low	Low	5
13	Low	Low	Good	High	Low	18
14	Low	High	Good	High	Low	15
15	Asbestos Content Unknown					48
16	Asbestos Found Present but Later Removed or Corrected or Asbestos Thought Not Present					881

a A school can only appear in one stratum.

Table 21. Distribution of New York City Public Schools by Selected Size Categories

School Enrollment	Number of Public Schools
1 - 50	2
51 - 100	16
101 - 200	30
201 - 300	29
301 - 400	45
401 - 500	54
501 - 750	232
751 - 1,000	212
1,001 +	378
Total	998

The information in this table was taken from the 1978 Curriculum Information Directory.

F. Construction and Stratification of the Second-Stage Frame

The second-stage frame will consist of all eligible sites in the first-stage sample of schools. A trained rater will visit each sample school and list all eligible sites. By visual inspection, the rater will rate each of these sites as to friability, condition, exposure, and accessibility. All sites on the second-stage frame will be stratified into the 16 categories formed from all possible low/high combinations of friability, condition, exposure, and accessibility. The 16 second-stage strata are listed in Table 22.

In this stratification proposed for the second-stage frame, the factor of asbestos content is not used. There are several reasons for this. First, the distribution of the first-stage sample of schools with respect to asbestos content has already been controlled to some extent by first-stage stratification. The use of asbestos content to stratify sites at the second stage presents several difficulties. Information on asbestos content is not readily available for all sites; in fact, in some schools bulk sampling was performed at only one site. This situation requires the following: (1) development of a method to classify as many sites as possible as to asbestos content, using the available information, and (2) application of a two-phase selection procedure to the sites that cannot be classified. Developing a classification method, (1), will undoubtedly result in some misclassifications and the number of these misclassifications will increase as the classification procedure becomes less strict. The two-phase selection procedure, (2), will involve bulk sampling at a sample of the sites that cannot be classified. This bulk sampling and laboratory analysis will take time and will be costly. Another aspect of the problem is that the information currently available on asbestos content is not in terms of percentages but is in categorical

Table 22. Second-Stage Strata

Stratum	Friability	Condition	Exposure	Accessibility
1	Low	Good	Low	Low
2	Low	Good	Low	High
3	Low	Good	High	Low
4	Low	Good	High	High
5	Low	Bad	Low	Low
6	Low	Bad	Low	High
7	Low	Bad	High	Low
8	Low	Bad	High	High
9	High	Good	Low	Low
10	High	Good	Low	High
11	High	Good	High	Low
12	High	Good	High	High
13	High	Bad	Low	Low
14	High	Bad	Low	High
15	High	Bad	High	Low
16	High	Bad	High	High

form. The categories used are (1) 1 to 5 percent, (2) 6 to 50 percent, and (3) more than 50 percent. Thus, 50 percent is used as the low/high asbestos content dividing point. It is thought more appropriate, on the basis of data reviewed, that a number in the neighborhood of 30 percent be used as the dividing point.

Considering all of the factors discussed above, it is concluded that not using asbestos content in second-stage stratification will save time and money, while most probably not substantially altering the distribution of the sample from that desired. Perhaps the cost savings at this point can be used to increase the total number of sites in the sample.

G. Allocation and Selection of the Second-Stage Sample

The total sample size will be allocated among the second-stage strata proportional to the stratum totals of number of sites (i.e., if stratum 2 has 10 percent of the sites, then approximately 10 percent of samples will be from strata 2), with the restriction that at least one site be selected from each nonempty second-stage stratum. Within second-stage strata, sites will be selected with equal probability and without replacement. It should be noted that, with the potential control of the distribution of the second-stage sample implied by a large number of strata at this stage, selection of the second-stage sample independently within first-stage units is not possible given the proposed allocation. As a result, sampling variances are not estimable unless the design is replicated. Therefore, to facilitate variance estimation, it is proposed to select two independent samples of sites (of equal size) from the second-stage frame.

As previously mentioned, it is likely that the first-stage sample size will be 16 schools. Possible total sample sizes are 32 sites, 48 sites, or 64 sites. (Cost/variance modelling to determine the optimal average number of sites

per school was not within the scope of this task.) Suppose the total sample size is 48 sites. Then two independent samples of 24 sites will be selected from the second-stage frame. The 24 sites will be allocated among the second-stage strata proportional to the stratum totals of number of sites, but selecting at least one site from each nonempty stratum. Intuitively speaking, proportional allocation is in response to the desire to obtain more information where more of the population of interest exists. For example, suppose that .001 percent of all sites in the study area have factor level combination 1, and 30 percent of all sites have factor level combination 2. Then it is more important, in terms of recommending future use of an algorithm, to validate the algorithm for factor level combination 2.

H. Data Collection at Sample Sites

At each sample site, algorithm factors will be rated, bulk samples of friable material will be collected, and air sampling will be performed. Both trained and untrained raters will rate the algorithm factors. Collection of bulk samples of friable material will follow the guidance given in Lucas et al. (1980), using the sample site as the Sampling Area. This guidance suggests that 3, 5, or 7 bulk samples be collected, depending on whether the size of the Sampling Area is less than 1,000 square feet, between 1,000 and 5,000 square feet, or greater than 5,000 square feet, respectively. The guidance given in Lucas et al. (1980) concerning laboratory analysis should also be followed. Air sampling will be conducted as described in Price et al. (1980b). A randomization procedure for placement of air samplers within sample sites is not within the scope of this study. However, an exact protocol for placement of a sampler within a site should be developed prior to air sampling. This protocol should provide standardization in the procedures used across all sites. It is very important that this be done in order to avoid jeopardizing the validity of the study results.

I. Design Effect

To insure that survey results are adequately precise to satisfy user needs, minimum allowable precision is usually specified, at least for the more pertinent estimates being sought. The sample variance or precision associated with parameter estimates depends upon the following characteristics of the sample design:

(a) the sample size to be used, depending upon the cost contraints associated with the study,
(b) the magnitude of the finite population effect, d_f,
(c) the magnitude of the stratification effect, d_s,
(d) the magnitude of the clustering effect, d_c,
(e) the magnitude of the unequal weighting effect, d_w.

The product of the factors mentioned in (b) through (e) above is referred to as the design effect (DEFF), as discussed in Kish (1965). The design effect is the ratio of the variance obtained using the sample design to the variance that would have been obtained using a simple random sample of the same size. The sample size (n) divided by the design effect (DEFF) is referred to as the effective sample size (n^*) of the sample design:

$$n^* = n/DEFF .$$

The effective sample size is the sample size that would be required under simple random sampling to obtain the precision resulting from using the sample design with sample size n.

In estimating the design effect expected under the sample design described in this section, the effects of clustering and unequal weighting are considered. The effect of the finite population correction factor is assumed to be negligible, as are gains in precision due to stratification. The clustering effect, d_c, is the increase in variance of the sample due to the homogeneity of clusters. This effect can be expressed as

$$d_c = 1 + \rho(\bar{n}_2 - 1) ,$$

where ρ is the intracluster correlation and $\bar{n}_2$ is the average cluster size, or the average number of sample sites per school. The intracluster correlation ρ measures the homogeneity within clusters in terms of the protion of the total element variance that is due to cluster membership. Possible values of $\bar{n}_2$ for this study are 2, 3, and 4 sites per school. An appropriate value of ρ appears not to be known with any degree of certainty, although values in the range $.1 \leq \rho \leq .3$ may be reasonable. The relationship of airborne asbestos levels in sites within the same school versus between schools may be affected by several factors. More than for sites in different schools, sites within the same school may have friable material with similar asbestos content and be similar in age and condition. Air circulation patterns within a school may also affect the relationship of airborne asbestos levels among sites within the same school. Taking these factors into account, intracluster correlations of .1, .2, and .3 are considered. Table 23 shows the clustering effects corresponding to these values of ρ and average cluster sizes of 2, 3, and 4.

The unequal weighting effect is the increase in variance due to the fact that the sampling weights (inverse selection probabilities) are unequal. The sampling weight for a site is constructed from (1) a first-stage component, school selection probability, and (2) a second-stage component, site selection probability, conditional on the first-stage sample of schools. Because this second-stage component depends on the first-stage sample of schools, the sampling weights are not known prior to sample selection. It is thought, however, that the major unequal weighting impact will occur at the first stage of selection. This is due to the following reasons: (1) a somewhat proportional allocation will be used at the second stage of selection, and (2) if the first-stage sample of schools is allocated equally

Table 23. The Clustering Effect Corresponding to Selected Values of ρ and $\bar{n}_2$

Average No. of Sites per School ($\bar{n}_2$)	Intracluster Correlation		
	.1	.2	.3
2	1.1	1.2	1.3
3	1.2	1.4	1.6
4	1.3	1.6	1.9

The entries in this table were calculated according to

$$\text{Clustering effect } d_c = 1 + \rho(\bar{n}_2 - 1),$$

where ρ is the intracluster correlation, and $\bar{n}_2$ is the average number of sites per school.

among the first-stage strata (see Table 20), then the second-stage strata (see Table 22) are not expected to vary greatly in number of sites (e.g., not over a tenfold difference), although some differences are certainly expected. Thus, only the first stage of selection will be considered in the unequal weighting effect calculations. Also, first-stage strata 15 and 16 (see Table 20) will not be included in these calculations as it is thought that the variance in airborne asbestos levels within each of these strata will be very small compared to that within each of the other 14 strata.

To facilitate calculation of the unequal weighting effect, it is assumed that schools are selected from first-stage strata with equal probability. (In fact, schools are to be selected with probability proportional to size measures based on enrollment.) Assuming that one school is selected from each stratum, the unequal weighting effect can be expressed as

$$d_w = 14 \sum_{i=1}^{14} N_i^2 \Big/ \left(\sum_{i=1}^{14} N_i \right)^2 ,$$

where N_i is the number of schools in first-stage stratum i. Taking the values of N_i from Table 20,

$$d_w = 14 \times 675/69^2$$

$$d_w = 1.985 \doteq 2.$$

Under all of the assumptions stated, it follows that the design effect is

$$DEFF = 2 \times d_c ,$$

where d_c is the clustering effect given in Table 23. A reasonable value of the intracluster correlation is .2. Then the design effect is expected to be 2 x 1.2 = 2.4 when the average number of sites per school is 2. This means that the ratio of the variance under this sample design to the variance using a simple random sample of the same size is

expected to be 2.4. When the average number of sites per school is 3, then the design effect is expected to be 2.8. The design effect is expected to be 3.2 when there is an average of 4 sites per school.

J. Sampling Strategy

Measurements of airborne asbestos concentration levels have been collected in schools and public buildings by several investigators (see Sebastien 1980, Nicholson 1975 and 1978, Logue 1981, Patton 1980). Unfortunately, correlations of these airborne levels with algorithm scoring has had limited success to date. This could be due to several factors including (1) the episodic nature of airborne concentrations which may be undetected with air sampling is carried out over relatively short time periods; (2) the problem of establishing background asbestos levels; and (3) the fact that previous studies were not specifically designed to investigate the relationship between airborne asbestos levels and algorithm scores over a wide range of conditions (e.g., high asbestos, low friability, good condition, etc.). The currently proposed field study has been designed so that several of these previous shortcomings have been taken into account (i.e., air sampling for one week at each site, background air levels at each site, and a designed experiment which covers a wide range of algorithm scores). Hopefully, this design will lead to a significant relationship between the various variables.

However, because of the limited success in past studies and the fact that air sampling is extremely expensive, it seems reasonable for the proposed study described in this report to be selective in which sites are initially sampled. That is, it is suggested that the initial 3 or 4 sites sampled have widely different algorithm scores (for example, two sites with high asbestos, friability, exposure, accessibility, and bad conditions (i.e., high algorithm scores), and two sites with low asbestos, friability, exposure, acces-

sibility, and good condition (i.e., low algorithm scores)). In conjunction with these initial sites, air sampling will also be done for background levels at each site. These initial comparisons (i.e., between high and low scoring sites and background sites) should indicate early in the study if there are going to be detectable relationships between air levels and algorithm scores. If significant relationships are not noted in these early sites, then a reassessment of the field study should be undertaken to determine if further sampling is worthwhile, considering the cost of asbestos air sampling.

REFERENCES

Britt HI, Luecke RH. May 1973. The estimation of parameters in nonlinear, implicit models. Technometrics, Vol. 15, No. 2.

EPA. September 1980a. Conference on the Exposure Assessment Algorithm. U.S. Environmental Protection Agency. Washington, D.C.

EPA. October 1980b. Asbestos-containing materials in schools: health effects and magnitude of exposure. U.S. Environmental Protection Agency.

Kish L. 1965. Survey Sampling. John Wiley & Sons, Inc. New York.

Logue EE, Hartwell TD. January 1981. Characteristics of the asbestos exposure algorithm: empirical distributions, correlations, and measurement validity. Research Triangle Institute. Research Triangle Park, North Carolina.

Lucas D, Hartwell TD, Rao AV. December 1980. Asbestos-containing materials in school buildings: guidance for asbestos analytical program. Research Triangle Institute. Prepared for U.S. Environmental Protection Agency. Washington, D.C. Contract No. EPA 560/13-80-017A.

Nicholson WJ, et al. 1975. Asbestos-contamination of the air in public buildings. Contract No. 68-02-1346. Environmental Prtection Agency, Office of Air and Waste Management, Office of Air Quality Planning and Standards: Environmental Sciences Laboratory: Mt. Sinai School of Medicine of the City University of New York.

Nicholson WJ, et al. 1978. Control of sprayed asbestos surfaces in school buildings: a feasibility study. Report to the National Institute of Environmental Health Sciences. Environmental Sciences Laboratory: Mt. Sinai School of Medicine of the City University of New York.

Owen DW. 1962. Handbook of statistical tables. Addison-Wesley.

Patton JL, et al. March 1980. Draft final report on asbestos in schools. Battelle Columbus Laboratories. Columbus, Ohio.

References (continued)

Price BP, Townley CW. November 1980a. Asbestos algorithm validation studies. Battelle Columbus Laboratories. Columbus, Ohio.

Price BP, Melton C, Schmidt E, Townley CW. November 1980b. Airborne asbestos levels in schools: a design study. Battelle Columbus Laboratories. Columbus, Ohio. Contract No. 68-01-3858 to U.S. Environmental Protection Agency, Washington, D.C.

Sebastien P, et al. August 1980. Measurement of asbestos air pollution inside buildings sprayed with asbestos. Laboratories d'Etude des Particules Inhalées. Paris, France.

APPENDIX A

The Noncentral t Distribution

APPENDIX A

The Noncentral t-Distribution

The tables of statistical power appearing in this report were developed using graphs of the noncentral t-distribution found in Handbook of Statistical Tables by D.B. Owen, Addison-Wesley, 1962. The critical parameter that must be identified in order to use the tables is the noncentrality parameter of the distribution. The noncentrality parameter, δ, takes different forms depending on the exact null hypothesis and alternative being tested. As a point of reference, we record the form of the noncentrality parameter used in each table. Where interpretation is necessary, some additional discussion is presented.

Table 3: $H_0: \mu_{10} = \mu_B$; $H_1: \mu_{10} = \mu_B + K\sigma$

$\delta = (n/2)^{1/2} K$

Table 7: same as Table 3

Table 10: $H_0: \beta_1 = 0$; $H_1: \beta_1 = K\sigma/\Delta$

$\delta = K \Sigma_S/\Delta$

where $\Sigma_S = \left(\sum (S_i - \bar{S})^2\right)^{1/2}$

Table 11: same as Table 10

Table 12: same as Table 10 with $\Delta = 30$.

Table 14: $H_0: \mu_A = \mu_{DA}$; $H_1: \mu_A = \mu_{DA} + K\sigma$

$\delta = m^{1/2} K / \left(\frac{1}{5} + \frac{1}{11}\right)^{1/2}$

To determine this probability, the noncentral t-distribution must be evaluated at $1.64 / (R\ \beta_1^2 + 1)^{1/2}$ for the various values of $R\ \beta_1^2$. These probabilities are not available from the tables in Owen (1962).

Table 15: $H_0: \mu_A = \mu_{DA}$; $H_1: \mu_A = \mu_{DA} + K\sigma$

$$\delta = \frac{m^{1/2} K}{\left(\frac{1}{5} + \frac{1}{11}\right)^{1/2}} \left[\frac{(1 - 2p)^2}{1 + K^2 p (1 - p)}\right]^{1/2}$$

The case is slightly more complicated where the hypotheses are $H_0: \beta_1 = 0$ and $H_1: \beta_1 = K\sigma/\Delta$ and it is assumed that algorithm scoring is not error free. In this case, it is not possible to obtain power directly from the published tables found in Owen. The difficulty is that the variance of the estimator of β_1 depends on the value of β_1 hypothesized under H_1. (Estimation of β_1 follows Britt and Luecke, 1973). The variance is given as

$$\text{Var } (\hat{\beta}_1) = (R \beta_1^2 + 1) \sigma_\varepsilon^2 / \sum_S{}^2$$

where $R = \sigma_\eta^2/\sigma_\varepsilon^2$ as defined in the text of the report. The statistical test of $H_0: \beta_1 = 0$ (significance level set at 5 percent) says to reject H_0 if

$$\frac{\hat{\beta}_1}{s_\varepsilon/\sum_S} > 1.64 .$$

To calculate the power of the test, we need to compute the probability that

$$\frac{\hat{\beta}_1}{s_\varepsilon/\sum_S} > 1.64$$

when β_1 is not zero. Appropriate algebraic manipulation leads to the following expression for power:

$$\text{Prob} \left\{ \text{noncentral t variate} > 1.64 / (R \beta_1^2 + 1)^{1/2} \right\}$$

with

$$\delta = \sum_S \beta_1/\sigma_\varepsilon \ (R \beta_1^2 + 1)^{1/2}.$$

To determine this probability, the noncentral t-distribution must be evaluated at $1.64/(R \beta_1{}^2 + 1)^{1/2}$ for various values of $R \beta_1{}^2$. These probabilities are not available from the tables in Owen (1962).